279	<800	2301	2337	655	<800
1323	1307	2753	2805	1341	1323
12	<800	490	<800	1563	1521
2101	2067	177	<800	690	<800
4290	4236	6836	6740	998	1005
1154	1166	6767	6740	509	<800
766	801	5595	5684		
6886	6882	5959	5945		
7307	7318	10600	10609		
		5616	5684		
		5062	5054		
		111	<800		
		4421	4433		
		1890	1892		
		4176	4156		

```
FEATURES             Location/Qualifiers
     source          1. .149940
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-56P22"
                     /clone_lib="RPCI human BAC library 11"

Query Match          13.0%;  Score 120.4;  DB 9;  Length 149940;
Best Local Similarity  61.5%;  Pred. No. 5.1e-19;
Matches  368;  Conservative    0;  Mismatches  156;  Indels   74;  Gaps    8;

Qy    317 TTTCTCCCCCAGTTACTTTCCAGCTCCGGGCTGGCTCAGGACCCGTGTTCCTCAGTGGCC 376
          ||||||| ||| | |||||  | ||| | ||| ||||   | || ||||| |||||| |
Db  79005 TTTCTCCTCCAATCACTTTTTAACTCTGAGCT-GCTCCAAATCCCTGTTCTTCAGTGACT 79063

Qy    377 AGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAGGAAGAAGAAGGGGAGG 436
          |   | |||||||| | | || |||||||||||| ||||| ||  | || | || ||| |
Db  79064 AAAGATGTTATGAAACTTTAGCCCTAACCTGGGAAGAGGAAGAAAAGGAGGGAGAGGAAG 79123

Qy    437 AGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATATATCTCTGGAGGAGC 496
          | ||||| |||| |||                      | |  | |||||  |||||
Db  79124 ACGAGGAGGAGGGAGA----------------------ATAGTTGTCTCTAAAGGAGG 79159

Qy    497 AAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCGTGGCTAAGA 556
          ||| | || ||||||| |||||| |   ||  ||| |||||| | | | |   |||||||
Db  79160 AAACCTCTATCAAACATGTCAAAGGTGGGGGACCCAAGAAGCTGAC-AACACTGCTAAGA 79218

Qy    557 AAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTAGAGACAAGAGCCCTG 616
          |  ||||||||||||          ||      |   ||| ||||||||||||||||
Db  79219 GCGAGGAGCTGGAAAAAG----------GATTGAGATCCATCTTAAAGACAAGAGCCCTG 79268

Qy    617 TGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAATGAGGAGCCA 676
          ||| ||||  |||| ||| |  || | | ||||| ||||| | |||||||| ||||||||
Db  79269 TGAGAAAGAACAAACCCATACTCAAACCTAAGAAACCAGGTTCCAAGAAATAAGGAGCCA 79328

Qy    677 ----------------------------CGCCTTGGGGGGCACGGTGCAAAGTGGGCC 706
                                       ||||  | |  ||   |||| |||||||||
Db  79329 GGAACTGCCTGACTGCGGTGATCAAGAGAGGGCCTGTGAGACCAGCATGCAGAGTGGGCC 79388

Qy    707 TTCCCTGGGCTGTGCTGCAGGCACAGGGTGCCCCTGTCC---AGCCCCTCCACCTGTGTC 763
          |||||||||||||  | ||  |||||  ||  | |     ||| |||| ||| |||
```

```
Db  79389 TTCCTTGGGCTGTGCTGTGGTCATGGGGTGTACCCCTACCTTGCCCCTTCCATCTGAGTC 79448

Qy    764 TGAATGCAACAGGGGTGTTGCGGGGGCAACATGAGA----GCCCCTCACCCCCAACTCTC 819
          | |||  ||  || |||||  || |  |     | |     | | || | | |||||||||
Db  79449 TAAATAGAATGGGAGTGTTTTGGAGAAAGTGACAAAAGATCCTCTTCTCACTCAACTCTC 79508

Qy    820 CACTTTCAGGAGGCCCCCA-GTGAAGAGCCCCACCTCGGGGTCACAATAAAGTTGCCT 876
          |  | | |||||||||     | ||||||||  ||| | |||  ||||||||||  ||
Db  79509 CCATGTGAGGAGGCCCTGGTGAGAAGAGCCTGACCACAGGGAAACAATAAAGTTTGCT 79566
```

RESULT 11
XLNUPLR

```
LOCUS       XLNUPLR                   646 bp    mRNA    linear   VRT 10-FEB-1999
DEFINITION  Xenopus laevis mRNA fragment for nucleoplasmin.
ACCESSION   Y00204
VERSION     Y00204.1  GI:64938
KEYWORDS    nucleoplasmin.
SOURCE      Xenopus laevis.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 646)
  AUTHORS   De Robertis,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-1987) de Robertis E., University of California,
            Dept. of Biological Chemistry, University of California, Los
            Angeles, CA 90024, USA
REFERENCE   2  (bases 1 to 646)
  AUTHORS   Burglin,T.R., Mattaj,I.W., Newmeyer,D.D., Zeller,R. and De
            Robertis,E.M.
  TITLE     Cloning of nucleoplasmin from Xenopus laevis oocytes and analysis
            of its developmental expression
  JOURNAL   Genes Dev. 1 (1), 97-107 (1987)
  MEDLINE   88112783
   PUBMED   3428591
FEATURES             Location/Qualifiers
     source          1. .646
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="N7.2"
                     /clone_lib="lambda gt11 cDNA library"
     CDS             <1. .575
                     /codon_start=3
                     /product="nucleoplasmin"
                     /protein_id="CAA68363.1"
                     /db_xref="GI:833629"
                     /db_xref="SWISS-PROT:P05221"
                     /translation="NTSKVEKPVSLIWGCELNEQNKTFAFKIEDEEEKCEHQLALRTV
                     CLGDKAKDEFHIVEIVTQEEGKEKPVPIASLKPSILPMATMVGIELTPPVTFRLKAGS
                     GPVYISGQHVAMEEDYSWAEEEDEGEEEEEEEEDPESPPKAVKRPAATKKAGQAKKKK
                     LDKEDESSEEDSPTKKGKGAGRGRKPAAKK"
     polyA_signal    634. .639
BASE COUNT       222 a    121 c    174 g    129 t
ORIGIN

Query Match          10.8%;  Score 99.6;  DB 5;  Length 646;
Best Local Similarity  57.4%;  Pred. No. 4.8e-14;
Matches  224;  Conservative    0;  Mismatches  154;  Indels   12;  Gaps    2;

Qy     69 GACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGACTTGGACCTTCAGACCCCA 128
          | || |  | ||||| || || || | | || |||    ||| |   | || | |   |
Db     29 GTCCCTTATTTGGGGGTGTGAACTGAATGAGCAGAACAAGACATTTGCGTTTAAAATAGA 88

Qy    129 GCTGGAGGGGAAGCAGAGCTGCAGGCTGTTG---CTTCATACGATTTGCTTGGGGGAGAA 185
          ||||   | | |       | ||||   || |  ||  | ||  | ||||| ||
Db     89 AGACGAGGAAGAAAAATGTGAGCATCAGTTGGCACTGCGCACAGTATGTCTTGGGGACAA 148

Qy    186 AGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCAGCAAACCAGGAGGACAAGAA 245
          || || || ||| | ||    || || ||          ||  |  || |||||    ||
Db    149 GGCAAAGGATGAGTTCCACATAGTAGAAAT---------AGTTACACAAGAGGAGGGAAA 199
```

```
Qy    246 GATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCCCATGGTCTCCATGGT 305
Db    200 GGAAAAACCAGTTCCGATTGCCAGTCTAAAGCCTTCTATTTTACCCATGGCAACCATGGT 259
Qy    306 AGGAGTGCAGCTTTCTCCCCCAGTTACTTTCCAGCTCCGGGCTGGCTCAGGACCCGTGTT 365
Db    260 GGGCATTGAGCTGACCCCTCCAGTTACTTTCCGGTTAAAAGCTGGTTCCGGCCCAGTGTA 319
Qy    366 CCTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAGGAAGA 425
Db    320 CATCAGTGGTCAACACGTAGCAATGGAGGAAGATTACTCATGGGCAGAAGAGGAAGATGA 379
Qy    426 AGAAGGGGAGGAGGAGGAAGAGGAAGAGGA 455
Db    380 AGGAGAAGAGGAGGAGGAAGAGGAAGAAGA 409
```

RESULT 12
XLNUPLR2
LOCUS XLNUPLR2 800 bp mRNA linear VRT 12-SEP-1993
DEFINITION Xenopus laevis mRNA for nucleoplasmin.
ACCESSION X04766
VERSION X04766.1 GI:64939
KEYWORDS nucleoplasmin.
SOURCE Xenopus laevis.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 800)
AUTHORS Dingwall,C., Dilworth,S.M., Black,S.J., Kearsey,S.E., Cox,L.S. and Laskey,R.A.
TITLE Nucleoplasmin cDNA sequence reveals polyglutamic acid tracts and a cluster of sequences homologous to putative nuclear localization signals
JOURNAL EMBO J. 6 (1), 69-74 (1987)
MEDLINE 87218476
PUBMED 2884102
FEATURES

Source	Location/Qualifiers
	1. .800
	/organism="Xenopus laevis"
	/db_xref="taxon:8355"
CDS	114. .716
	/note="nucleoplasmin (AA 1-200)"
	/codon_start=1
	/protein_id="CAA28460.1"
	/db_xref="GI:64940"
	/db_xref="SWISS-PROT:P05221"
	/translation="MASTVSNTSKLEKPVSLIWGCELNEQDKTFEFKVEDDEEKCEHQ LALRTVCLGDKAKDEFNIVEIVTQEEGAEKSVPIATLKPSILPMATMVGIELTPPVTF RLKAGSGPLYISGQHVAMEEDYSWAEEEDEGEAEGEEEEEEEEDQESPPKAVKRPAAT KKAGQAKKKKLDKEDESSEEDSPTKKGKGAGRGRKPAAKK"
polyA_site	800
	/note="polyA site"

BASE COUNT 250 a 153 c 222 g 175 t
ORIGIN

Query Match 10.4%; Score 96.2; DB 5; Length 800;
Best Local Similarity 54.9%; Pred. No. 3.4e-13;
Matches 284; Conservative 0; Mismatches 188; Indels 45; Gaps 3;

```
Qy    153 GCTGTTGCTTCATACGATTTGCTTGGGGGAGAAAGCCAAAGAGGAGATGCATCGCGTGGA 212
Db    245 GTTGGCGTTGCGCACGGTGTGTCTGGGGGACAAGGCAAAGGATGAGTTCAACATTGTAGA 304
Qy    213 GATCCTGCCCCCAGCAAACCAGGAGGACAAGAAGATGCAGCCGGTCACCATTGCCTCACT 272
Db    305 AATCGTTACACAAGAGGAGGGAGCGGAAA---------AATCTGTTCCAATTGCCACTCT 355
Qy    273 CCAGGCCTCAGTCCTCCCCATGGTCTCCATGGTAGGAGTGCAGCTTTCTCCCCCAGTTAC 332
Db    356 AAAGCCTTCTATTCTACCCATGGCAACTATGGTGGGCATTGAGCTGACTCCTCCAGTTAC 415
```

```
Qy    333 TTTCCAGCTCCGGGCTGGCTCAGGACCCGTGTTCCTCAGTGGCCAGGAACGTTATGAAGC 392
Db    416 TTTCCGGTTAAAAGCTGGTTCCGGCCCACTGTACATCAGTGGTCA---------------- 460
Qy    393 ATCAGACCTAACCTGGGAGGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGA 452
Db    461 ---------ACACGTAGCGATGGAGGAAGATTACTCATGGGCAGAAGAGGAAGATGAAGG 511
Qy    453 GGAAGATGATGAGGATGAGGATGCAGATATATCTCTGGAGGAGCAAAGCCCTGTCAAACA 512
Db    512 GGAAGCTGAAGGAGAAGAGGAGGAAGAAGAGGAAGAAGATCAAGAATCTCCACCCAAAGC 571
Qy    513 AGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAGCTGGAAAA 572
Db    572 TGTAAAGAGGCCTGCGGCTACCAAAAAAGCAGGCCAGGCAAAGAAGAAGAAACTTGACAA 631
Qy    573 AGAAGAAGAGGAAATAAGAGCCAGCGTTAGAGACAAGAGCCCTGTGAAAAAGGCCAAAGC 632
Db    632 AGAGGACGAG----------AGCTCCGAGGAAGACAGTCCAACCAAAAAGGGCAAAGG 679
Qy    633 CACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAATGA 669
Db    680 AGCCGGAAGAGGAAGAAAGCCGGCTGCTAAGAAGTGA 716
```

RESULT 13
I66494/c
LOCUS I66494 7218 bp DNA linear PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES

source	Location/Qualifiers
	1. .7218
	/organism="unknown"

BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match 10.1%; Score 93.2; DB 6; Length 7218;
Best Local Similarity 3.5%; Pred. No. 2.2e-12;
Matches 14; Conservative 257; Mismatches 125; Indels 0; Gaps 0;

```
Qy    364 TTCCTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAGGAA 423
Db   1441 TTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1382
Qy    424 GAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATATA 483
Db   1381 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1322
Qy    484 TCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCG 543
Db   1321 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1262
Qy    544 AGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTAGA 603
Db   1261 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1202
Qy    604 GACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAG 663
Db   1201 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1142
Qy    664 AAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTG 723
Db   1141 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1082
```

```
Best Local Similarity   3.5%;  Pred. No. 1.5e-14;
Matches   14;  Conservative  257;  Mismatches  125;  Indels    0;  Gaps    0;

Qy     364 TTCCTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAGGAA 423
           ||  |  : ::  ::::: :  : :::: :  :::   ::   :::::::::::::::::
Db    1441 TTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1382

Qy     424 GAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATATA 483
           :::::::::::::::::::::::::::::::::::: :: ::::: ::::: : ::: : :
Db    1381 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1322

Qy     484 TCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCG 543
           :::::::: ::::    :  ::: :::  ::::::  :: :   ::::: ::: :
Db    1321 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1262

Qy     544 AGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTAGA 603
           :: : ::  ::::::::::::  :::::::::::::::::::: ::::::  :: :  :::
Db    1261 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1202

Qy     604 GACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAG 663
           :: :::::    : :::::::::  ::::  : ::  ::::  ::::::  ::::   :::
Db    1201 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1142

Qy     664 AAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTG 723
           ::: ::::::  : :   :::::: : :: : :::: :::         :::  : :  :
Db    1141 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1082

Qy     724 CAGGCACAGGGTGCCCCTGTCCAGCCCCTCCACCTG 759
           ::: : :::: :     |      ||||| |||||
Db    1081 RRRRRRRRRRRRRRRATCGCAAGCTCCCTCGACCTG 1046
```

RESULT 2
US-08-781-891-209/c
; Sequence 209, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6090620tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 240052.419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 209:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51259 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-781-891-209

```
Query Match             6.9%;  Score 63.4;  DB 3;  Length 51259;
Best Local Similarity   59.9%;  Pred. No. 1.6e-06;
Matches  106;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy     408 GGAGGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGA 467
           ||||||||||||||| || || | |||||||||||| |||||  |||| || || |||||
Db     257 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAGGAGGAGGAGGAGGA 198

Qy     468 TGAGGATGCAGATATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGT 527
           |||||  |  ||          |||||||||      |   |  |     |  |  ||
Db     197 AGAGGAGGAGGAGGAGAAGAAGGAGGAGCAGGAGGAGGAGGAGGAGCAGGAGGAGGAGGA 138

Qy     528 GCCCCAGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGA 584
           |    || || ||| ||  | ||  |||  |      ||  |||  | || || |||||
Db     137 GAAGGAGGAGAAGGAGAAGGAGGAGAAGGAGGGGCAGGAGGAGGAGGAGGAGGAGGA 81
```

RESULT 3
US-08-781-891-208/c
; Sequence 208, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6090620tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 240052.419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 208:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16442 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-781-891-208

```
Query Match             6.4%;  Score 59.4;  DB 3;  Length 16442;
Best Local Similarity   57.6%;  Pred. No. 1.1e-05;
Matches  125;  Conservative    0;  Mismatches   91;  Indels    1;  Gaps    1;

Qy     376 CAGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAGGAAGAAGAAGGGGAG 435
           |||||  |   | || |     || |       | ||||||||||||| ||  | | ||||
Db   16369 CAGGAGCAGGAGGAGGAGCAGGAGCAGGAGGAGCAGGAGGAGGAGGAGGAGCAGGAGGAG 16310

Qy     436 GAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATATATCTCTGGAGGAG 495
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 24, 2003, 05:06:35 ; Search time 77 Seconds
(without alignments)
3680.121 Million cell updates/sec

Title: US-09-844-864-16
Perfect score: 924
Sequence: 1 cagcccgcttctctgcccgg.....tttgcggccgcaagcttatg 924

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	93.2	10.1	7218	1	US-08-232-463-14	Sequence 14, Appl
c	2	63.4	6.9	51259	3	US-08-781-891-209	Sequence 209, App
c	3	59.4	6.4	16442	3	US-08-781-891-208	Sequence 208, App
	4	59.2	6.4	3489	2	US-08-728-323A-1	Sequence 1, Appli
	5	59.2	6.4	3489	4	US-09-298-568-1	Sequence 1, Appli
c	6	59.2	6.4	32207	2	US-08-770-379-20	Sequence 20, Appl
c	7	59.2	6.4	32207	4	US-08-757-669A-20	Sequence 20, Appl
c	8	59.2	6.4	32207	4	US-09-230-371A-20	Sequence 20, Appl
	9	57.6	6.2	3211	2	US-08-574-959A-8	Sequence 8, Appli
	10	57.6	6.2	3211	4	US-09-357-014-8	Sequence 8, Appli
	11	57.6	6.2	3901	2	US-08-574-959A-6	Sequence 6, Appli
	12	57.6	6.2	3901	4	US-09-357-014-6	Sequence 6, Appli
	13	56.6	6.1	2518	4	US-09-433-699-3	Sequence 3, Appli
	14	55.2	6.0	2340	3	US-09-022-983-4	Sequence 4, Appli
	15	55.2	6.0	2477	4	US-09-490-692-3	Sequence 3, Appli
	16	54.4	5.9	9636	1	US-08-323-170B-1	Sequence 1, Appli
	17	54.4	5.9	9636	4	US-08-954-441-1	Sequence 1, Appli
	18	54	5.8	1236	2	US-08-741-134-5	Sequence 5, Appli
	19	52.2	5.6	289	4	US-09-007-005-17	Sequence 17, Appl
	20	52.2	5.6	289	4	US-09-244-796-17	Sequence 17, Appl
	21	52.2	5.6	15378	3	US-08-785-420-1	Sequence 1, Appli
	22	51.8	5.6	966	2	US-08-766-738-2	Sequence 2, Appli
	23	51.8	5.6	966	4	US-09-262-610-2	Sequence 2, Appli
	24	51.6	5.6	489	1	US-07-879-685B-3	Sequence 3, Appli
	25	50.6	5.5	5183	1	US-08-459-568-3	Sequence 3, Appli
	26	50.6	5.5	5183	2	US-08-399-411-3	Sequence 3, Appli
	27	50.6	5.5	5868	3	US-08-516-859A-3	Sequence 3, Appli
	28	50.6	5.5	5868	4	US-09-586-472-3	Sequence 3, Appli
	29	50.6	5.5	5868	4	US-09-528-706-3	Sequence 3, Appli
c	30	50.4	5.5	1926	4	US-09-249-585A-4	Sequence 4, Appli
c	31	50.4	5.5	1931	2	US-09-130-114-2	Sequence 2, Appli
	32	50	5.4	258	4	US-09-345-882-21	Sequence 21, Appl
	33	50	5.4	6002	4	US-09-345-882-4	Sequence 4, Appli
	34	50	5.4	162450	4	US-09-345-882-1	Sequence 1, Appli
	35	49.4	5.3	3100	1	US-08-296-362-1	Sequence 1, Appli
c	36	49	5.3	2188	1	US-07-865-662F-10	Sequence 10, Appl
c	37	49	5.3	2188	4	US-08-374-219B-10	Sequence 10, Appl
	38	48.8	5.3	1678	3	US-08-650-766-2	Sequence 2, Appli
	39	48.8	5.3	1954	3	US-08-922-635-2	Sequence 2, Appli
	40	48.8	5.3	2255	2	US-08-741-134-1	Sequence 1, Appli
	41	48.8	5.3	3318	3	US-08-650-766-3	Sequence 3, Appli
	42	48.8	5.3	3318	3	US-08-922-635-3	Sequence 3, Appli
	43	48.8	5.3	3385	3	US-08-650-766-1	Sequence 1, Appli
	44	48.8	5.3	3385	3	US-08-922-635-1	Sequence 1, Appli
	45	48.8	5.3	15202	3	US-08-922-635-21	Sequence 21, Appl

ALIGNMENTS

```
RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match                 10.1%;  Score 93.2;  DB 1;  Length 7218;
```

```
  Best Local Similarity   3.5%;  Pred. No. 1.5e-14;
  Matches   14;  Conservative  257;  Mismatches  125;  Indels    0;  Gaps    0;

Qy      364 TTCCTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAGGAA 423
            ||  |  : ::  ::::: :  : :::: :  :::   ::   ::::::::::::::::
Db     1441 TTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1382

Qy      424 GAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATATA 483
            :::::::::::::::::::::::::::::::::: :: ::::: ::::: : ::: : :
Db     1381 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1322

Qy      484 TCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCG 543
                :::::: :::    :  ::: :::  ::::::  :: :    ::::: ::: :
Db     1321 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1262

Qy      544 AGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTAGA 603
            :: : ::  :::::::::::  :::::::::::::::::::: :::::  :: :  :::
Db     1261 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1202

Qy      604 GACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAG 663
            :: :::::    : ::::::::  ::::  : ::  ::::  ::::::  ::::   :::
Db     1201 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1142

Qy      664 AAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTG 723
            ::: ::::::  : :    :::::: : :: : :::: :::         :::  : :  :
Db     1141 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1082

Qy      724 CAGGCACAGGGTGCCCCTGTCCAGCCCCTCCACCTG 759
             ::: : :::: :     |      ||||| |||||
Db     1081 RRRRRRRRRRRRRRRATCGCAAGCTCCCTCGACCTG 1046


RESULT 2
US-08-781-891-209/c
; Sequence 209, Application US/08781891
; Patent No. 6090620
;  GENERAL INFORMATION:
;    APPLICANT:  Fu, Ying-Hui
;    APPLICANT:  Yu, Chang-En
;    APPLICANT:  Oshima, Junko
;    APPLICANT:  Mulligan, John T.
;    APPLICANT:  Schellenberg, Gerald D.
;    TITLE OF INVENTION:  GENE AND GENE PRODUCTS RELATED TO
;    TITLE OF INVENTION:  WERNER'S SYNDROME
;    NUMBER OF SEQUENCES:  209
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,891
;      FILING DATE:  27-DEC-1996
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6090620tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  240052.419
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  209:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  51259 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-781-891-209

  Query Match             6.9%;  Score 63.4;  DB 3;  Length 51259;
  Best Local Similarity   59.9%;  Pred. No. 1.6e-06;
  Matches  106;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy      408 GGAGGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGA 467
            |||||||||||||| || || | |||||||||||| |||||  |||| || || |||||
Db      257 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAGGAGGAGGAGGAGGA 198

Qy      468 TGAGGATGCAGATATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGT 527
             ||||| |  ||         ||||||||       |   |  |     |   |  ||
Db      197 AGAGGAGGAGGAGGAGAAGAAGGAGGAGCAGGAGGAGGAGGAGGAGCAGGAGGAGGAGGA 138

Qy      528 GCCCCAGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGA 584
            |    || || ||| ||  | ||  ||| |     ||  |||  | || || |||||
Db      137 GAAGGAGGAGAAGGAGAAGGAGGAGAAGGAGGGGCAGGAGGAGGAGGAGGAGGAGGA 81


RESULT 3
US-08-781-891-208/c
; Sequence 208, Application US/08781891
; Patent No. 6090620
;  GENERAL INFORMATION:
;    APPLICANT:  Fu, Ying-Hui
;    APPLICANT:  Yu, Chang-En
;    APPLICANT:  Oshima, Junko
;    APPLICANT:  Mulligan, John T.
;    APPLICANT:  Schellenberg, Gerald D.
;    TITLE OF INVENTION:  GENE AND GENE PRODUCTS RELATED TO
;    TITLE OF INVENTION:  WERNER'S SYNDROME
;    NUMBER OF SEQUENCES:  209
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,891
;      FILING DATE:  27-DEC-1996
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6090620tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  240052.419
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  208:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  16442 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-781-891-208

  Query Match             6.4%;  Score 59.4;  DB 3;  Length 16442;
  Best Local Similarity   57.6%;  Pred. No. 1.1e-05;
  Matches  125;  Conservative    0;  Mismatches   91;  Indels    1;  Gaps    1;

Qy      376 CAGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAGGAAGAAGAAGGGGAG 435
            ||||| |   | || |     || |       | ||||||||||||| ||  | | ||||
Db    16369 CAGGAGCAGGAGGAGGAGCAGGAGCAGGAGGAGCAGGAGGAGGAGGAGGAGCAGGAGGAG 16310

Qy      436 GAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATATATCTCTGGAGGAG 495
```

```
Db  16309 CAGGAGGAGGAGGAGGAGCAGGAGGAGGAGCAGGAGCAGGAGGAGGAGGAGCAGGAGGAG 16250

Qy    496 CAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCGTGGCTAAG 555
Db  16249 CAGGAGGAGGAGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGTAGAAGGAGAAG 16190

Qy    556 AAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAG 592
Db  16189 AAGAAGAAG-AAGAAGAAGAAGAAGAAGAAGAAGAAG 16154
```

RESULT 4
US-08-728-323A-1
; Sequence 1, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3489
US-08-728-323A-1

Query Match 6.4%; Score 59.2; DB 2; Length 3489;
Best Local Similarity 52.8%; Pred. No. 6.2e-06;
Matches 151; Conservative 0; Mismatches 133; Indels 2; Gaps 1;

```
Qy    408 GGAGGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGA 467
Db   1101 GGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGA 1160

Qy    468 TGAGGATGCAGATATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGT 527
Db   1161 GGAGGACGAGGAGGAGGACGAGGAGGAGGACGAGGAGGAGGAGGACGAGGAGGATGACGA 1220

Qy    528 GCCCCAGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAAT 587
Db   1221 TGATGAGGACAATGAGGACGAGGAGGATGACGAGGAGGAGGACAAGAAGGAGGACGAGGA 1280

Qy    588 AAGAGCCAGCGTTAGAGACAAGAGCCCTGTGA--AAAAGGCCAAAGCCACAGCCAGAGCC 645
Db   1281 GGACGGGGGCGATGGAAACAAAACGTTGAGCATCCAAAGTTCACAACAGCAGCAGGAGCC 1340

Qy    646 AAGAAGCCAGGATTCAAGAAATGAGGAGCCACGCCTTGGGGGGCAC 691
Db   1341 ACAACAGCAGGAGCCACAGCAGCAGGAGCCACAGCAGCAGGAGCCC 1386
```

RESULT 5
US-09-298-568-1
; Sequence 1, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D.
; APPLICANT: Ballestas, Mary E.
; APPLICANT: Kaye, Kenneth M.
; TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
; TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
; FILE REFERENCE: 16412-10001R
; CURRENT APPLICATION NUMBER: US/09/298,568
; CURRENT FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,422
; EARLIER FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-1

Query Match 6.4%; Score 59.2; DB 4; Length 3489;
Best Local Similarity 52.8%; Pred. No. 6.2e-06;
Matches 151; Conservative 0; Mismatches 133; Indels 2; Gaps 1;

```
Qy    408 GGAGGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGA 467
Db   1101 GGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGA 1160

Qy    468 TGAGGATGCAGATATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGT 527
Db   1161 GGAGGACGAGGAGGAGGACGAGGAGGAGGACGAGGAGGAGGAGGACGAGGAGGATGACGA 1220

Qy    528 GCCCCAGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAAT 587
Db   1221 TGATGAGGACAATGAGGACGAGGAGGATGACGAGGAGGAGGACAAGAAGGAGGACGAGGA 1280

Qy    588 AAGAGCCAGCGTTAGAGACAAGAGCCCTGTGA--AAAAGGCCAAAGCCACAGCCAGAGCC 645
Db   1281 GGACGGGGGCGATGGAAACAAAACGTTGAGCATCCAAAGTTCACAACAGCAGCAGGAGCC 1340

Qy    646 AAGAAGCCAGGATTCAAGAAATGAGGAGCCACGCCTTGGGGGGCAC 691
Db   1341 ACAACAGCAGGAGCCACAGCAGCAGGAGCCACAGCAGCAGGAGCCC 1386
```

RESULT 6
US-08-770-379-20/c
; Sequence 20, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
; TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF

```
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/770,379
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  52342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32207 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-770-379-20

  Query Match             6.4%;  Score 59.2;  DB 2;  Length 32207;
  Best Local Similarity   52.8%;  Pred. No. 1.6e-05;
  Matches  151;  Conservative    0;  Mismatches  133;  Indels    2;  Gaps    1;

Qy     408 GGAGGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGA 467
           ||| || |||||||| || || | ||| || ||||| || || ||||| ||||| |||||
Db   20896 GGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGA 20837

Qy     468 TGAGGATGCAGATATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGT 527
           ||||| |  ||          ||||||| |       |   | | |   |       |
Db   20836 GGAGGACGAGGAGGAGGACGAGGAGGAGGACGAGGAGGAGGAGGACGAGGAGGATGACGA 20777

Qy     528 GCCCCAGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAAT 587
           || |  | | |  || ||   |  |  |  ||  ||| ||  | || || ||
Db   20776 TGATGAGGACAATGAGGACGAGGAGGATGACGAGGAGGAGGACAAGAAGGAGGACGAGGA 20717

Qy     588 AAGAGCCAGCGTTAGAGACAAGAGCCCTGTGA--AAAAGGCCAAAGCCACAGCCAGAGCC 645
           |   ||| | || |||| |         |   ||||  || | |  ||||  |||||
Db   20716 GGACGGGGGCGATGGAAACAAAACGTTGAGCATCCAAAGTTCACAACAGCAGCAGGAGCC 20657

Qy     646 AAGAAGCCAGGATTCAAGAAATGAGGAGCCACGCCTTGGGGGGCAC 691
           |  |   |||||  ||    |  |||||||||  |    || || |
Db   20656 ACAACAGCAGGAGCCACAGCAGCAGGAGCCACAGCAGCAGGAGCCC 20611


RESULT 7
US-08-757-669A-20/c
; Sequence 20, Application US/08757669A
; Patent No. 6183751
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Yuan
;    APPLICANT:  Bohenzky, Roy A.
;    APPLICANT:  Russo, James J.
;    APPLICANT:  Edelman, Isidore S.
;    APPLICANT:  Moore, Patrick S.
;    TITLE OF INVENTION:  UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
;    TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/757,669A
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  45185-F
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32207 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-757-669A-20

  Query Match             6.4%;  Score 59.2;  DB 4;  Length 32207;
  Best Local Similarity   52.8%;  Pred. No. 1.6e-05;
  Matches  151;  Conservative    0;  Mismatches  133;  Indels    2;  Gaps    1;

Qy     408 GGAGGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGA 467
           ||| || |||||||| || || | ||| || ||||| || || ||||| ||||| |||||
Db   20896 GGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGA 20837

Qy     468 TGAGGATGCAGATATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGT 527
           |||||  |  ||          ||||||| |       |   | | |   |       |
Db   20836 GGAGGACGAGGAGGAGGACGAGGAGGAGGACGAGGAGGAGGAGGACGAGGAGGATGACGA 20777

Qy     528 GCCCCAGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAAT 587
           || |  | | |  || ||   |  |  |  ||  ||| ||  | || || ||
Db   20776 TGATGAGGACAATGAGGACGAGGAGGATGACGAGGAGGAGGACAAGAAGGAGGACGAGGA 20717

Qy     588 AAGAGCCAGCGTTAGAGACAAGAGCCCTGTGA--AAAAGGCCAAAGCCACAGCCAGAGCC 645
           |   ||| | || |||| |         |   ||||  || | |  ||||  |||||
Db   20716 GGACGGGGGCGATGGAAACAAAACGTTGAGCATCCAAAGTTCACAACAGCAGCAGGAGCC 20657

Qy     646 AAGAAGCCAGGATTCAAGAAATGAGGAGCCACGCCTTGGGGGGCAC 691
           |  |   |||||  ||    |  |||||||||  |    || || |
Db   20656 ACAACAGCAGGAGCCACAGCAGCAGGAGCCACAGCAGCAGGAGCCC 20611


RESULT 8
US-09-230-371A-20/c
; Sequence 20, Application US/09230371A
; Patent No. 6348586
; GENERAL INFORMATION:
;  APPLICANT:  Chang, Yuan
;  APPLICANT:  Bohenzky, Roy A
;  APPLICANT:  Russo, James J
;  APPLICANT:  Edelman, Isidore S
;  APPLICANT:  Moore, Patrick S
;  TITLE OF INVENTION:  UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE:  45185-G-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/230,371A
;  CURRENT FILING DATE:  1999-11-17
;  PRIOR APPLICATION NUMBER:  PCT/US97/13346
;  PRIOR FILING DATE:  1997-07-22
;  NUMBER OF SEQ ID NOS:  30
```

```
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 20
;    LENGTH: 32207
;    TYPE: DNA
;    ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-20

  Query Match              6.4%;  Score 59.2;  DB 4;  Length 32207;
  Best Local Similarity   52.8%;  Pred. No. 1.6e-05;
  Matches  151;  Conservative    0;  Mismatches  133;  Indels    2;  Gaps    1;

Qy     408 GGAGGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGA 467
           ||| || |||||||| || || | ||| || ||||| || || ||||| ||||| |||||
Db   20896 GGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGA 20837

Qy     468 TGAGGATGCAGATATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGT 527
           ||||| |  ||         ||||||| |       |   | | |   |       |
Db   20836 GGAGGACGAGGAGGAGGACGAGGAGGAGGACGAGGAGGAGGAGGACGAGGAGGATGACGA 20777

Qy     528 GCCCCAGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAAT 587
             || |  | | |  || ||   |  |  |  ||  ||| ||  | || || ||
Db   20776 TGATGAGGACAATGAGGACGAGGAGGATGACGAGGAGGAGGACAAGAAGGAGGACGAGGA 20717

Qy     588 AAGAGCCAGCGTTAGAGACAAGAGCCCTGTGA--AAAAGGCCAAAGCCACAGCCAGAGCC 645
           |   ||| | || |||| |         |   ||||  || | |  |||| |||||
Db   20716 GGACGGGGGCGATGGAAACAAAACGTTGAGCATCCAAAGTTCACAACAGCAGCAGGAGCC 20657

Qy     646 AAGAAGCCAGGATTCAAGAAATGAGGAGCCACGCCTTGGGGGGCAC 691
           |  |   |||||  ||    |  |||||||||  |    || || |
Db   20656 ACAACAGCAGGAGCCACAGCAGCAGGAGCCACAGCAGCAGGAGCCC 20611


RESULT 9
US-08-574-959A-8
; Sequence 8, Application US/08574959A
; Patent No. 5962224
;  GENERAL INFORMATION:
;    APPLICANT:  Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi
;    APPLICANT:  and Jack L. Strominger
;    TITLE OF INVENTION:  p62 POLYPEPTIDES, RELATED POLYPEPTIDES
;    TITLE OF INVENTION:  AND USES THEREFOR
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, Suite 510
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/574,959A
;      FILING DATE:  19-DEC-95
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandragouras, Amy E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  DFN-008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3211 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  439..3157
US-08-574-959A-8

  Query Match              6.2%;  Score 57.6;  DB 2;  Length 3211;
  Best Local Similarity   57.1%;  Pred. No. 1.6e-05;
  Matches  105;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy     409 GAGGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGAT 468
           || |||||||||||||||| ||  ||||| || ||||| ||||| ||||| || || ||
Db    2398 GAAGAGGAGGAGGAAGAAGGAGAAGAGGAAGAAGAAGAAGAAGAAGAAGAGGAAGAA 2457

Qy     469 GAGGATGCAGATATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTG 528
           || || |  ||    |       |||||| ||  |  ||   || | |   |       | |
Db    2458 GAAGAAGAGGAAGAAGAGGAAGAGGAGGAAGACTTTGAGGAAGAGGAAGAGGATGAAGAG 2517

Qy     529 CCCCAGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATA 588
             |      | | |    |  |   || || || || | ||  |||||   ||||||  |
Db    2518 GAATATTTTGAAGAGGAAGAAGAGGAGGAAGAAGAGTTTGAGGAAGAATTTGAGGAAGAA 2577

Qy     589 AGAG 592
           ||
Db    2578 GAAG 2581


RESULT 10
US-09-357-014-8
; Sequence 8, Application US/09357014
; Patent No. 6291645
;  GENERAL INFORMATION:
;       APPLICANT: Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi
;                  and Jack L. Strominger
;       TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
;                           AND USES THEREFOR
;       NUMBER OF SEQUENCES: 22
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: LAHIVE & COCKFIELD
;            STREET: 60 State Street, Suite 510
;            CITY: Boston
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02109-1875
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/357,014
;            FILING DATE: 19-Jul-1999
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/574,959
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Mandragouras, Amy E.
;            REGISTRATION NUMBER: 36,207
;            REFERENCE/DOCKET NUMBER: DFN-008
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (617)227-7400
;            TELEFAX: (617)227-5941
;  INFORMATION FOR SEQ ID NO: 8:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 3211 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION:  439..3157
;       SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-357-014-8

  Query Match              6.2%;  Score 57.6;  DB 4;  Length 3211;
```

```
  Best Local Similarity   57.1%;  Pred. No. 1.6e-05;
  Matches  105;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy     409 GAGGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGAT 468
           || |||||||||||||||| ||  ||||| || ||||| ||||| ||||| || || ||
Db    2398 GAAGAGGAGGAGGAAGAAGGAGAAGAGGAAGAAGAAGAAGAAGAAGAAGAGGAAGAA 2457

Qy     469 GAGGATGCAGATATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTG 528
           || || |  ||   |      |||||| ||  |  ||   || | |   |       | |
Db    2458 GAAGAAGAGGAAGAAGAGGAAGAGGAGGAAGACTTTGAGGAAGAGGAAGAGGATGAAGAG 2517

Qy     529 CCCCAGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATA 588
               |     | | |   |  |   || || || || | ||  |||||   ||||||  |
Db    2518 GAATATTTTGAAGAGGAAGAAGAGGAGGAAGAAGAGTTTGAGGAAGAATTTGAGGAAGAA 2577

Qy     589 AGAG 592
             ||
Db    2578 GAAG 2581


RESULT 11
US-08-574-959A-6
; Sequence 6, Application US/08574959A
; Patent No. 5962224
;  GENERAL INFORMATION:
;    APPLICANT:  Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi
;    APPLICANT:  and Jack L. Strominger
;    TITLE OF INVENTION:  p62 POLYPEPTIDES, RELATED POLYPEPTIDES
;    TITLE OF INVENTION:  AND USES THEREFOR
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, Suite 510
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/574,959A
;      FILING DATE:  19-DEC-95
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandragouras, Amy E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  DFN-008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3901 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  439..3847
US-08-574-959A-6

  Query Match             6.2%;  Score 57.6;  DB 2;  Length 3901;
  Best Local Similarity   57.1%;  Pred. No. 1.7e-05;
  Matches  105;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy     409 GAGGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGAT 468
           || |||||||||||||||| ||  ||||| || ||||| ||||| ||||| || || ||
Db    3088 GAAGAGGAGGAGGAAGAAGGAGAAGAGGAAGAAGAAGAAGAAGAAGAAGAGGAAGAA 3147

Qy     469 GAGGATGCAGATATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTG 528
           || || |  ||   |      |||||| ||  |  ||   || | |   |       | |
Db    3148 GAAGAAGAGGAAGAAGAGGAAGAGGAGGAAGACTTTGAGGAAGAGGAAGAGGATGAAGAG 3207

Qy     529 CCCCAGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATA 588
               |     | | |   |  |   || || || || | ||  |||||   ||||||  |
Db    3208 GAATATTTTGAAGAGGAAGAAGAGGAGGAAGAAGAGTTTGAGGAAGAATTTGAGGAAGAA 3267

Qy     589 AGAG 592
             ||
Db    3268 GAAG 3271


RESULT 12
US-09-357-014-6
; Sequence 6, Application US/09357014
; Patent No. 6291645
;   GENERAL INFORMATION:
;        APPLICANT: Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi
;                   and Jack L. Strominger
;        TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
;                            AND USES THEREFOR
;        NUMBER OF SEQUENCES: 22
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: LAHIVE & COCKFIELD
;             STREET: 60 State Street, Suite 510
;             CITY: Boston
;             STATE: Massachusetts
;             COUNTRY: USA
;             ZIP: 02109-1875
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/357,014
;             FILING DATE: 19-Jul-1999
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/574,959
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Mandragouras, Amy E.
;             REGISTRATION NUMBER: 36,207
;             REFERENCE/DOCKET NUMBER: DFN-008
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (617)227-7400
;             TELEFAX: (617)227-5941
;   INFORMATION FOR SEQ ID NO: 6:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 3901 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  439..3847
;        SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-357-014-6

  Query Match             6.2%;  Score 57.6;  DB 4;  Length 3901;
  Best Local Similarity   57.1%;  Pred. No. 1.7e-05;
  Matches  105;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy     409 GAGGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGAT 468
           || |||||||||||||||| ||  ||||| || ||||| ||||| ||||| || || ||
Db    3088 GAAGAGGAGGAGGAAGAAGGAGAAGAGGAAGAAGAAGAAGAAGAAGAAGAGGAAGAA 3147

Qy     469 GAGGATGCAGATATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTG 528
           || || |  ||   |      |||||| ||  |  ||   || | |   |       | |
Db    3148 GAAGAAGAGGAAGAAGAGGAAGAGGAGGAAGACTTTGAGGAAGAGGAAGAGGATGAAGAG 3207

Qy     529 CCCCAGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATA 588
```

```
Db    3208 GAATATTTTGAAGAGGAAGAAGAGGAGGAAGAAGAGTTTGAGGAAGAATTTGAGGAAGAA 3267

Qy     589 AGAG 592
           ||
Db    3268 GAAG 3271


RESULT 13
US-09-433-699-3
; Sequence 3, Application US/09433699B
; Patent No. 6165786
; GENERAL INFORMATION:
;  APPLICANT: C. Frank Bennett
;  APPLICANT: Lex M. Cowsert
;  TITLE OF INVENTION: ANTISENSE MODULATION OF NUCLEOLIN EXPRESSION
;  FILE REFERENCE: RTS-0109
;  CURRENT APPLICATION NUMBER: US/09/433,699B
;  CURRENT FILING DATE: 1999-11-03
;  NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 3
;   LENGTH: 2518
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (112)..(2235)
US-09-433-699-3

  Query Match             6.1%;  Score 56.6;  DB 4;  Length 2518;
  Best Local Similarity   53.4%;  Pred. No. 2.5e-05;
  Matches  119;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy     357 ACCCGTGTTCCTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGA 416
Db     795 AGCCAAGAACGTGGCTGAGGATGAAGATGAAGAAGAGGATGATGAGGACGAGGATGACGA 854

Qy     417 GGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGC 476
Db     855 CGACGACGAAGATGATGAAGATGATGATGATGAAGATGATGAGGAGGAGGAAGAAGAGGA 914

Qy     477 AGATATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAA 536
Db     915 GGAGGAAGAGCCTGTCAAAGAAGCACCTGGAAAACGAAAGAAGGAAATGGCCAAACAGAA 974

Qy     537 GCAGGCGAGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAA 579
Db     975 AGCAGCTCCTGAAGCCAAGAAACAGAAAGTGGAAGGCACAGAA 1017


RESULT 14
US-09-022-983-4
; Sequence 4, Application US/09022983
; Patent No. 6159731
;  GENERAL INFORMATION:
;    APPLICANT:  Yang, Xiaolu
;    APPLICANT:  Khosravi-Far, Roya
;    APPLICANT:  Chang, Howard Y.
;    APPLICANT:  Baltimore, David
;    TITLE OF INVENTION:  DAXX, A NOVEL FAS-BINDING
;    TITLE OF INVENTION:  PROTEIN THAT ACTIVATES JNK AND APOPTOSIS
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wolf, Greenfield & Sacks, P.C.
;      STREET:  600 Atlantic Avenue
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/022,983
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/037,919
;      FILING DATE:  12-FEB-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/051,753
;      FILING DATE:  26-JUN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Van Amsterdam, John R.
;      REGISTRATION NUMBER:  40,212
;      REFERENCE/DOCKET NUMBER:  M0656/7036
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-720-3500
;      TELEFAX:  617-720-2441
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2340 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  1...2220
US-09-022-983-4

  Query Match             6.0%;  Score 55.2;  DB 3;  Length 2340;
  Best Local Similarity   55.1%;  Pred. No. 5.7e-05;
  Matches  108;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy     423 AGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATAT 482
Db    1296 AGCTGAGACAGATGACGAAGACGATGAGGAGAGTGATGAGGAAGAGGAGGAGGAGGAGGA 1355

Qy     483 ATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGC 542
Db    1356 AGAAGAAGAGGAGGAGGCCACAGATTCTGAAGAGGAGGAGGATCTGGAACAGATGCAGGA 1415

Qy     543 GAGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTAG 602
Db    1416 GGGTCAGGAGGATGATGAAGAGGAGGACGAAGAGGAAGAAGCAGCAGCAGGTAAAGATGG 1475

Qy     603 AGACAAGAGCCCTGTG 618
           ||||||||||||  ||
Db    1476 AGACAAGAGCCCCATG 1491


RESULT 15
US-09-490-692-3
; Sequence 3, Application US/09490692
; Patent No. 6180353
; GENERAL INFORMATION:
;  APPLICANT: Nicholas M. Dean
;  APPLICANT: Lex M. Cowsert
;  TITLE OF INVENTION: ANTISENSE MODULATION OF DAXX EXPRESSION
;  FILE REFERENCE: RTS-0120
;  CURRENT APPLICATION NUMBER: US/09/490,692
;  CURRENT FILING DATE: 2000-01-24
;  NUMBER OF SEQ ID NOS: 176
; SEQ ID NO 3
;   LENGTH: 2477
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (116)...(2338)
US-09-490-692-3
```

Query Match 6.0%; Score 55.2; DB 4; Length 2477;
Best Local Similarity 55.1%; Pred. No. 5.8e-05;
Matches 108; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```
Qy     423 AGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATAT 482
           ||  ||    || || || || || || |||   ||||| || || ||||| |  ||
Db    1411 AGCTGAGACAGATGACGAAGACGATGAGGAGAGTGATGAGGAAGAGGAGGAGGAGGAGA 1470

Qy     483 ATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGC 542
           |      |||||| |   | | |      |||   |   |    ||   |||| |||||
Db    1471 AGAAGAAGAGGAGGAGGCCACAGATTCTGAAGAGGAGGAGGATCTGGAACAGATGCAGGA 1530

Qy     543 GAGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTAG 602
           | |   ||   |  |  || ||  |||  |||| ||||| | |  |  ||  |  | | |
Db    1531 GGGTCAGGAGGATGATGAAGAGGAGGACGAAGAGGAAGAAGCAGCAGCAGGTAAAGATGG 1590

Qy     603 AGACAAGAGCCCTGTG 618
           ||||||||||||  ||
Db    1591 AGACAAGAGCCCCATG 1606
```

Search completed: January 24, 2003, 18:46:33
Job time : 206 secs

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 23, 2003, 14:06:24 ; Search time 45576 Seconds
(without alignments)
590.025 Million cell updates/sec

Title: US-09-844-864-16
Perfect score: 924
Sequence: 1 cagcccgcttctctgcccgg.....tttgcggccgcaagcttatg 924

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	883	95.6	1874	9	AK094267	AK094267 Homo sapi
	2	337	36.5	423	6	AX321879	AX321879 Sequence
	3	262.4	28.4	156347	2	AC091171	AC091171 Homo sapi
	4	262.4	28.4	166878	2	AC023288	AC023288 Homo sapi
	5	262.4	28.4	181714	2	AC090442	AC090442 Homo sapi
	6	262.4	28.4	184050	2	AC087819	AC087819 Homo sapi
c	7	262.4	28.4	184444	2	AC090471	AC090471 Homo sapi
c	8	171.6	18.6	86574	9	HS833B7	AL008637 Human DNA
	9	122	13.2	135206	2	AC108486	AC108486 Homo sapi
	10	120.4	13.0	149940	9	AC093557	AC093557 Homo sapi
	11	99.6	10.8	646	5	XLNUPLR	Y00204 Xenopus lae
	12	96.2	10.4	800	5	XLNUPLR2	X04766 Xenopus lae
c	13	93.2	10.1	7218	6	I66494	I66494 Sequence 14
	14	90.6	9.8	198421	2	AC125180	AC125180 Mus muscu
	15	83.2	9.0	190604	2	AC122303	AC122303 Mus muscu
	16	81.6	8.8	200052	2	AL772398	AL772398 Mus muscu
c	17	81.4	8.8	251076	2	AC097410	AC097410 Rattus no
	18	81.2	8.8	2081	9	AK056804	AK056804 Homo sapi
c	19	79.8	8.6	43553	9	AL357372	AL357372 Human DNA
	20	79	8.5	222259	2	AC113078	AC113078 Mus muscu
	21	78.4	8.5	241666	2	AC118543	AC118543 Mus muscu
	22	77.4	8.4	182740	2	AC115811	AC115811 Mus muscu
c	23	76.8	8.3	165077	10	AC084382	AC084382 Mus muscu
	24	76.4	8.3	170032	2	AC116673	AC116673 Mus muscu
c	25	76.2	8.2	70391	2	AC110410	AC110410 Rattus no
c	26	75.8	8.2	144328	9	AC009695	AC009695 Homo sapi
	27	75.8	8.2	163495	9	AC022716	AC022716 Homo sapi
	28	75.8	8.2	170631	9	AC020751	AC020751 Homo sapi
	29	75.8	8.2	170807	9	AC023812	AC023812 Homo sapi
	30	75.8	8.2	179803	9	AC068314	AC068314 Homo sapi
	31	75.4	8.2	168856	2	AF466883	AF466883 Mus muscu
c	32	75.4	8.2	169153	2	AC128501	AC128501 Rattus no
	33	75.2	8.1	110000	2	AC098456_2	Continuation (3 of
	34	75.2	8.1	303943	2	AC127311	AC127311 Mus muscu
	35	74.8	8.1	177716	9	AC026188	AC026188 Homo sapi
c	36	74.6	8.1	152763	2	AC119715	AC119715 Rattus no
c	37	74.6	8.1	175456	2	AC115183	AC115183 Rattus no
c	38	74.4	8.1	147670	10	AC084020	AC084020 Mus muscu
	39	74.4	8.1	205606	10	AL596204	AL596204 Mouse DNA
c	40	74	8.0	184370	2	AL589845	AL589845 Mus muscu
c	41	73.8	8.0	120044	2	AC129042	AC129042 Rattus no
c	42	73.8	8.0	208844	2	AC125070	AC125070 Mus muscu
	43	73.8	8.0	215938	2	AC102646	AC102646 Mus muscu
c	44	73.6	8.0	93682	10	AL591805	AL591805 Mouse DNA
c	45	73.6	8.0	154733	2	AC131221	AC131221 Rattus no

ALIGNMENTS

RESULT 1
AK094267
LOCUS AK094267 1874 bp mRNA linear PRI 15-JUL-2002
DEFINITION Homo sapiens cDNA FLJ36948 fis, clone BRACE2005719, weakly similar to NUCLEOPLASMIN.
ACCESSION AK094267
VERSION AK094267.1 GI:21753294
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens cerebellum cDNA to mRNA, clone_lib:BRACE2 clone:BRACE2005719.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Tanigami,A., Fujiwara,T., Shibahara,T., Goto,Y., Hirao,M.,

Shimizu,F., Wakebe,H., Ono,T., Hishigaki,H., Watanabe,T., Ozaki,K., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1874)
AUTHORS Isogai,T. and Yamamoto,J.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.
FEATURES Location/Qualifiers
source 1. .1874
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="BRACE2005719"
/tissue_type="cerebellum"
/clone_lib="BRACE2"
/note="cloning vector: pME18SFL3"
BASE COUNT 394 a 570 c 604 g 306 t
ORIGIN

Query Match 95.6%; Score 883; DB 9; Length 1874;
Best Local Similarity 100.0%; Pred. No. 3.2e-208;
Matches 883; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CAGCCCGCTTCTCTGCCCGGAGCCATGAATCTCAGTAGCGCCAGTAGCACGGAGGAAAAG 60
Db     992 CAGCCCGCTTCTCTGCCCGGAGCCATGAATCTCAGTAGCGCCAGTAGCACGGAGGAAAAG 1051

Qy      61 GCAGTGACGACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGACTTGGACCTTC 120
Db    1052 GCAGTGACGACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGACTTGGACCTTC 1111

Qy     121 AGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCATACGATTTGCTTGGGG 180
Db    1112 AGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCATACGATTTGCTTGGGG 1171

Qy     181 GAGAAAGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCAGCAAACCAGGAGGAC 240
Db    1172 GAGAAAGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCAGCAAACCAGGAGGAC 1231

Qy     241 AAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCCCATGGTCTCC 300
Db    1232 AAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCCCATGGTCTCC 1291

Qy     301 ATGGTAGGAGTGCAGCTTTCTCCCCCAGTTACTTTCCAGCTCCGGGCTGGCTCAGGACCC 360
Db    1292 ATGGTAGGAGTGCAGCTTTCTCCCCCAGTTACTTTCCAGCTCCGGGCTGGCTCAGGACCC 1351

Qy     361 GTGTTCCTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAG 420
Db    1352 GTGTTCCTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAG 1411

Qy     421 GAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGAT 480
Db    1412 GAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGAT 1471

Qy     481 ATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAG 540
Db    1472 ATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAG 1531

Qy     541 GCGAGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTT 600
Db    1532 GCGAGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTT 1591

Qy     601 AGAGACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTC 660
Db    1592 AGAGACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTC 1651

Qy     661 AAGAAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTG 720
Db    1652 AAGAAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTG 1711

Qy     721 CTGCAGGCACAGGGTGCCCCTGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTG 780
Db    1712 CTGCAGGCACAGGGTGCCCCTGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTG 1771

Qy     781 TTGCGGGGGCAACATGAGAGCCCCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGT 840
Db    1772 TTGCGGGGGCAACATGAGAGCCCCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGT 1831

Qy     841 GAAGAGCCCCACCTCGGGGTCACAATAAAGTTGCCTGGTCAGG 883
Db    1832 GAAGAGCCCCACCTCGGGGTCACAATAAAGTTGCCTGGTCAGG 1874
```

RESULT 2
AX321879
LOCUS AX321879 423 bp DNA linear PAT 15-DEC-2001
DEFINITION Sequence 410 from Patent WO0172295.
ACCESSION AX321879
VERSION AX321879.1 GI:17906455
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Reed,S.G., Lodes,M.J., Mohamath,R., Secrist,H., Benson,D.R., Indirias,C.Y., Henderson,R.A., Fling,S.P., Algate,P.A., Elliot,M., Mannion,J. and Kalos,M.D.
TITLE Compositions and methods for the therapy and diagnosis of lung cancer
JOURNAL Patent: WO 0172295-A 410 04-OCT-2001;
CORIXA CORPORATION (US)
FEATURES Location/Qualifiers
source 1. .423
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 84 a 138 c 125 g 76 t
ORIGIN

Query Match 36.5%; Score 337; DB 6; Length 423;
Best Local Similarity 100.0%; Pred. No. 6.5e-73;
Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CAGCCCGCTTCTCTGCCCGGAGCCATGAATCTCAGTAGCGCCAGTAGCACGGAGGAAAAG 60
Db      87 CAGCCCGCTTCTCTGCCCGGAGCCATGAATCTCAGTAGCGCCAGTAGCACGGAGGAAAAG 146

Qy      61 GCAGTGACGACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGACTTGGACCTTC 120
Db     147 GCAGTGACGACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGACTTGGACCTTC 206

Qy     121 AGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCATACGATTTGCTTGGGG 180
Db     207 AGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCATACGATTTGCTTGGGG 266

Qy     181 GAGAAAGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCAGCAAACCAGGAGGAC 240
Db     267 GAGAAAGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCAGCAAACCAGGAGGAC 326

Qy     241 AAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCCCATGGTCTCC 300
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    327 AAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCCCATGGTCTCC 386

Qy    301 ATGGTAGGAGTGCAGCTTTCTCCCCCAGTTACTTTCC 337
          |||||||||||||||||||||||||||||||||||||
Db    387 ATGGTAGGAGTGCAGCTTTCTCCCCCAGTTACTTTCC 423


RESULT 3
AC091171
LOCUS       AC091171               156347 bp    DNA     linear   HTG 23-AUG-2002
DEFINITION  Homo sapiens chromosome 8 clone RP11-67H12 map 8, *** SEQUENCING IN
            PROGRESS ***, 6 unordered pieces.
ACCESSION   AC091171
VERSION     AC091171.7  GI:22296939
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156347)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-67H12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156347)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 156347)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 18, 2002 this sequence version replaced gi:22123594.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12352
                Center clone name: 67_H_12
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    44565: contig of 44565 bp in length
            *    44566 44665: gap of      100 bp
            *    44666    99572: contig of 54907 bp in length
            *    99573 99672: gap of      100 bp
            *    99673   104341: contig of 4669 bp in length
            *   104342 104441: gap of      100 bp
            *   104442   123153: contig of 18712 bp in length
            *   123154 123253: gap of      100 bp
            *   123254   137735: contig of 14482 bp in length
            *   137736 137835: gap of      100 bp
            *   137836   156347: contig of 18512 bp in length.
FEATURES             Location/Qualifiers
     source          1. .156347
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-67H12"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    39805 a  37308 c  37797 g  40934 t    503 others
ORIGIN

  Query Match           28.4%;  Score 262.4;  DB 2;  Length 156347;
  Best Local Similarity  99.6%;  Pred. No. 3.2e-54;
  Matches  263;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     621 AAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAATGAGGAGCCACGCC 680
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   86968 ACAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAATGAGGAGCCACGCC 87027

Qy     681 TTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTGCAGGCACAGGGTGCCCC 740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   87028 TTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTGCAGGCACAGGGTGCCCC 87087

Qy     741 TGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTGTTGCGGGGGCAACATGAGAG 800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   87088 TGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTGTTGCGGGGGCAACATGAGAG 87147

Qy     801 CCCCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGTGAAGAGCCCCACCTCGGGGT 860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   87148 CCCCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGTGAAGAGCCCCACCTCGGGGT 87207

Qy     861 CACAATAAAGTTGCCTGGTCAGGA 884
           ||||||||||||||||||||||||
Db   87208 CACAATAAAGTTGCCTGGTCAGGA 87231


RESULT 4
AC023288
LOCUS       AC023288               166878 bp    DNA     linear   HTG 22-FEB-2001
DEFINITION  Homo sapiens chromosome 8 clone RP11-67H12, WORKING DRAFT SEQUENCE,
            15 unordered pieces.
ACCESSION   AC023288
VERSION     AC023288.11  GI:9280725
```

```
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 166878)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Lam,B., Marathe,R., Miranda,M., Morehouse,A.J., Nguyen,M.,
            Oefner,P., Palm,C.J., Ramirez,D., Southwick,A.M., Wilhelmy,J.,
            Yu,S. and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 166878)
  AUTHORS   Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D., Federspiel,N.,
            Glukhov,S., Hansen,N., Hyman,R., Mao,J., Marathe,R.,
            Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D., Wilhelmy,J.,
            Yu,S. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-FEB-2000) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Jul 19, 2000 this sequence version replaced gi:8980905.
            -------------- Genome Center
                 Center: Stanford DNA Sequencing and Technology Development
            Center
                 Center code: SDSTDC
                 Web site: http://sequence-www.stanford.edu/group/human/
                 Contact: hum-info@sequence.stanford.edu
            -------------- Project Information
                 Center project name: 836
                 Center clone name: RP11-67H12
            -------------- Summary Statistics
                 Sequencing Vector: M13mp18; X02513
                 Chemistry: Dye-primer; 0% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 151238 bases at least Q40
                 Consensus quality: 155909 bases at least Q30
                 Consensus quality: 157935 bases at least Q20
                 Insert size: 116494; agarose-fp
                 Insert size: 165478; sum-of-contigs
                 Quality coverage:  9.4x in Q20 bases; agarose-fp
                 Quality coverage:  6.6x in Q20 bases; sum-of-contigs.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2175: contig of 2175 bp in length
            *     2176     2275: gap of unknown length
            *     2276     4472: contig of 2197 bp in length
            *     4473     4572: gap of unknown length
            *     4573     6489: contig of 1917 bp in length
            *     6490     6589: gap of unknown length
            *     6590     8363: contig of 1774 bp in length
            *     8364     8463: gap of unknown length
            *     8464    10644: contig of 2181 bp in length
            *    10645    10744: gap of unknown length
            *    10745    13373: contig of 2629 bp in length
            *    13374    13473: gap of unknown length
            *    13474    17686: contig of 4213 bp in length
            *    17687    17786: gap of unknown length
            *    17787    26030: contig of 8244 bp in length
            *    26031    26130: gap of unknown length
            *    26131    33532: contig of 7402 bp in length
            *    33533    33632: gap of unknown length
            *    33633    46267: contig of 12635 bp in length
            *    46268    46367: gap of unknown length
            *    46368    61796: contig of 15429 bp in length
            *    61797    61896: gap of unknown length
            *    61897    78516: contig of 16620 bp in length
            *    78517    78616: gap of unknown length
            *    78617    97532: contig of 18916 bp in length
            *    97533    97632: gap of unknown length
            *    97633   132687: contig of 35055 bp in length
            *   132688   132787: gap of unknown length
            *   132788   166878: contig of 34091 bp in length.
FEATURES             Location/Qualifiers
     source          1. .166878
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="RP11-67H12"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1. .2175
                     /note="assembly_name:Contig22"
     misc_feature    2276. .4472
                     /note="assembly_name:Contig24"
     misc_feature    4573. .6489
                     /note="assembly_name:Contig25"
     misc_feature    6590. .8363
                     /note="assembly_name:Contig28"
     misc_feature    8464. .10644
                     /note="assembly_name:Contig31"
     misc_feature    10745. .13373
                     /note="assembly_name:Contig33"
     misc_feature    13474. .17686
                     /note="assembly_name:Contig34"
     misc_feature    17787. .26030
                     /note="assembly_name:Contig35"
     misc_feature    26131. .33532
                     /note="assembly_name:Contig36
                     clone_end:SP6
                     vector_side:left"
     misc_feature    33633. .46267
                     /note="assembly_name:Contig37"
     misc_feature    46368. .61796
                     /note="assembly_name:Contig38"
     misc_feature    61897. .78516
                     /note="assembly_name:Contig39"
     misc_feature    78617. .97532
                     /note="assembly_name:Contig40
                     clone_end:T7
                     vector_side:right"
     misc_feature    97633. .132687
                     /note="assembly_name:Contig41"
     misc_feature    132788. .166878
                     /note="assembly_name:Contig42"
BASE COUNT    41842 a  40044 c  40620 g  42959 t   1413 others
ORIGIN

  Query Match             28.4%;  Score 262.4;  DB 2;  Length 166878;
  Best Local Similarity   99.6%;  Pred. No. 3.2e-54;
  Matches  263;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     621 AAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAATGAGGAGCCACGCC 680
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 146754 ACAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAATGAGGAGCCACGCC 146813

Qy     681 TTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTGCAGGCACAGGGTGCCCC 740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 146814 TTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTGCAGGCACAGGGTGCCCC 146873

Qy     741 TGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTGTTGCGGGGGCAACATGAGAG 800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 146874 TGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTGTTGCGGGGGCAACATGAGAG 146933

Qy     801 CCCCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGTGAAGAGCCCCACCTCGGGGT 860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 146934 CCCCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGTGAAGAGCCCCACCTCGGGGT 146993

Qy     861 CACAATAAAGTTGCCTGGTCAGGA 884
           ||||||||||||||||||||||||
```

```
Db 146994 CACAATAAAGTTGCCTGGTCAGGA 147017


RESULT 5
AC090442
LOCUS       AC090442               181714 bp    DNA     linear   HTG 26-JUL-2002
DEFINITION  Homo sapiens chromosome 8 clone RP11-868P18, WORKING DRAFT
            SEQUENCE, 15 unordered pieces.
ACCESSION   AC090442
VERSION     AC090442.2  GI:14389319
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_CANCELLED.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 181714)
  AUTHORS   Reichwald,K., Baumgart,C., Blechschmidt,K., Dette,M., Jahn,N.,
            Lehmann,R., Menzel,U., Polley,A., Schilhabel,M.B., Schudy,A.,
            Siddiqui,R., Taudien,S., Wen,G., Rosenthal,A. and Platzer,M.
  TITLE     Chromosome 8 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 181714)
  AUTHORS   Genome Sequencing Center Jena.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-2001) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On Jun 13, 2001 this sequence version replaced gi:13112136.
            ------------- Genome Center
                 Center: Insitute of Molecular Biotechnoloy
                 Center code: IMB
                 Web site: http://genome.imb-jena.de/
                 Contact: gscj-submit@genome.imb-jena.de
            ------------- Project Information
                 Center project name: H522
                 Center clone name: RP11-868P18
            ------------- Summary Statistics
                 Sequencing vector: M13; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990329
                 Consensus quality: 167442 bases at least Q40
                 Consensus quality: 173537 bases at least Q30
                 Consensus quality: 179477 bases at least Q20
                 Quality coverage: 5.19 x in Q20 bases; sum-of-contigs
            --------------.
             Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality 10.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
            --------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3436: contig of 3436 bp in length
            *     3437     3536: gap of unknown length
            *     3537    10120: contig of 6584 bp in length
            *    10121    10220: gap of unknown length
            *    10221    12191: contig of 1971 bp in length
            *    12192    12291: gap of unknown length
            *    12292    22694: contig of 10403 bp in length
            *    22695    22794: gap of unknown length
            *    22795    42595: contig of 19801 bp in length
            *    42596    42695: gap of unknown length
            *    42696    43816: contig of 1121 bp in length
            *    43817    43916: gap of unknown length
            *    43917    60529: contig of 16613 bp in length
            *    60530    60629: gap of unknown length
            *    60630    66757: contig of 6128 bp in length
            *    66758    66857: gap of unknown length
            *    66858    71578: contig of 4721 bp in length
            *    71579    71678: gap of unknown length
            *    71679    76238: contig of 4560 bp in length
            *    76239    76338: gap of unknown length
            *    76339    88520: contig of 12182 bp in length
            *    88521    88620: gap of unknown length
            *    88621   105835: contig of 17215 bp in length
            *   105836   105935: gap of unknown length
            *   105936   117374: contig of 11439 bp in length
            *   117375   117474: gap of unknown length
            *   117475   163683: contig of 46209 bp in length
            *   163684   163783: gap of unknown length
            *   163784   181714: contig of 17931 bp in length.
FEATURES             Location/Qualifiers
     source          1. .181714
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="RP11-868P18"
     misc_feature    1. .4
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    163784. .181714
                     /note="Please note that a sequence read was obtained using
                     vectorprimers. This read represents the right end of the
                     clone. However, within this read, the cloning site, e.g.
                     the exact end of insert could not be identified.'"
BASE COUNT    48155 a  40820 c  40243 g  51096 t   1400 others
ORIGIN

  Query Match             28.4%;  Score 262.4;  DB 2;  Length 181714;
  Best Local Similarity   99.6%;  Pred. No. 3.2e-54;
  Matches  263;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      621 AAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAATGAGGAGCCACGCC 680
            | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    35142 ACAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAATGAGGAGCCACGCC 35201

Qy      681 TTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTGCAGGCACAGGGTGCCCC 740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    35202 TTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTGCAGGCACAGGGTGCCCC 35261

Qy      741 TGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTGTTGCGGGGGCAACATGAGAG 800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    35262 TGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTGTTGCGGGGGCAACATGAGAG 35321

Qy      801 CCCCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGTGAAGAGCCCCACCTCGGGGT 860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    35322 CCCCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGTGAAGAGCCCCACCTCGGGGT 35381

Qy      861 CACAATAAAGTTGCCTGGTCAGGA 884
            ||||||||||||||||||||||||
Db    35382 CACAATAAAGTTGCCTGGTCAGGA 35405


RESULT 6
AC087819
LOCUS       AC087819               184050 bp    DNA     linear   HTG 09-MAY-2001
DEFINITION  Homo sapiens chromosome 8 clone RP11-507M15 map 8, WORKING DRAFT
            SEQUENCE, 20 unordered pieces.
ACCESSION   AC087819
VERSION     AC087819.2  GI:13194968
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 184050)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-507M15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 184050)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 4, 2001 this sequence version replaced gi:12584317.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11344
                Center clone name: 507_M_15
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 175465 bases at least Q40
                Consensus quality: 179683 bases at least Q30
                Consensus quality: 181216 bases at least Q20
                Insert size: 170000; agarose-fp
                Insert size: 182150; sum-of-contigs
                Quality coverage: 7.1 in Q20 bases; agarose-fp
                Quality coverage: 6.7 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 20 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    29338: contig of 29338 bp in length
            *    29339 29438: gap of      100 bp
            *    29439    29457: contig of 19 bp in length
            *    29458 29557: gap of      100 bp
            *    29558    30186: contig of 629 bp in length
            *    30187 30286: gap of      100 bp
            *    30287    31083: contig of 797 bp in length
            *    31084 31183: gap of      100 bp
            *    31184    31978: contig of 795 bp in length
            *    31979 32078: gap of      100 bp
            *    32079    32748: contig of 670 bp in length
            *    32749 32848: gap of      100 bp
            *    32849    34030: contig of 1182 bp in length
            *    34031 34130: gap of      100 bp
            *    34131    35157: contig of 1027 bp in length
            *    35158 35257: gap of      100 bp
            *    35258    72407: contig of 37150 bp in length
            *    72408 72507: gap of      100 bp
            *    72508    75697: contig of 3190 bp in length
            *    75698 75797: gap of      100 bp
            *    75798    78108: contig of 2311 bp in length
            *    78109 78208: gap of      100 bp
            *    78209    82150: contig of 3942 bp in length
            *    82151 82250: gap of      100 bp
            *    82251    88573: contig of 6323 bp in length
            *    88574 88673: gap of      100 bp
            *    88674    93173: contig of 4500 bp in length
            *    93174 93273: gap of      100 bp
            *    93274   101047: contig of 7774 bp in length
            *   101048 101147: gap of      100 bp
            *   101148   111284: contig of 10137 bp in length
            *   111285 111384: gap of      100 bp
            *   111385   130797: contig of 19413 bp in length
            *   130798 130897: gap of      100 bp
            *   130898   152534: contig of 21637 bp in length
            *   152535 152634: gap of      100 bp
            *   152635   182885: contig of 30251 bp in length
            *   182886 182985: gap of      100 bp
            *   182986   184050: contig of 1065 bp in length.
FEATURES             Location/Qualifiers
     source          1. .184050
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-507M15"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .29338
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    29439. .29457
                     /note="assembly_fragment"
     misc_feature    29558. .30186
                     /note="assembly_fragment"
     misc_feature    30287. .31083
                     /note="assembly_fragment"
     misc_feature    31184. .31978
                     /note="assembly_fragment"
     misc_feature    32079. .32748
                     /note="assembly_fragment"
     misc_feature    32849. .34030
                     /note="assembly_fragment"
     misc_feature    34131. .35157
                     /note="assembly_fragment"
     misc_feature    35258. .72407
                     /note="assembly_fragment"
     misc_feature    72508. .75697
                     /note="assembly_fragment"
     misc_feature    75798. .78108
                     /note="assembly_fragment"
     misc_feature    78209. .82150
                     /note="assembly_fragment"
     misc_feature    82251. .88573
                     /note="assembly_fragment"
     misc_feature    88674. .93173
                     /note="assembly_fragment"
     misc_feature    93274. .101047
                     /note="assembly_fragment"
     misc_feature    101148. .111284
                     /note="assembly_fragment"
     misc_feature    111385. .130797
                     /note="assembly_fragment"
     misc_feature    130898. .152534
                     /note="assembly_fragment"
```

```
     misc_feature    152635. .182885
                     /note="assembly_fragment"
     misc_feature    182986. .184050
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    49622 a  42339 c  42431 g  47728 t   1930 others
ORIGIN

  Query Match             28.4%;  Score 262.4;  DB 2;  Length 184050;
  Best Local Similarity   99.6%;  Pred. No. 3.2e-54;
  Matches  263;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      621 AAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAATGAGGAGCCACGCC 680
            | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   123140 ACAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAATGAGGAGCCACGCC 123199

Qy      681 TTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTGCAGGCACAGGGTGCCCC 740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   123200 TTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTGCAGGCACAGGGTGCCCC 123259

Qy      741 TGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTGTTGCGGGGGCAACATGAGAG 800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   123260 TGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTGTTGCGGGGGCAACATGAGAG 123319

Qy      801 CCCCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGTGAAGAGCCCCACCTCGGGGT 860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   123320 CCCCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGTGAAGAGCCCCACCTCGGGGT 123379

Qy      861 CACAATAAAGTTGCCTGGTCAGGA 884
            |||||||||||||||||||||||||
Db   123380 CACAATAAAGTTGCCTGGTCAGGA 123403


RESULT 7
AC090471/c
LOCUS       AC090471               184444 bp    DNA     linear   HTG 30-MAY-2001
DEFINITION  Homo sapiens chromosome 8 clone RP11-868P18 map 8, WORKING DRAFT
            SEQUENCE, 10 unordered pieces.
ACCESSION   AC090471
VERSION     AC090471.6  GI:14249070
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 184444)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-868P18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 184444)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 30, 2001 this sequence version replaced gi:14196391.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11354
                Center clone name: 868_P_18
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 180698 bases at least Q40
                Consensus quality: 182469 bases at least Q30
                Consensus quality: 183168 bases at least Q20
                Insert size: 183000; agarose-fp
                Insert size: 183544; sum-of-contigs
                Quality coverage: 15.1 in Q20 bases; agarose-fp
                Quality coverage: 15.1 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    24646: contig of 24646 bp in length
            *    24647 24746: gap of      100 bp
            *    24747    28060: contig of 3314 bp in length
            *    28061 28160: gap of      100 bp
            *    28161    33796: contig of 5636 bp in length
            *    33797 33896: gap of      100 bp
            *    33897    41745: contig of 7849 bp in length
            *    41746 41845: gap of      100 bp
            *    41846    52788: contig of 10943 bp in length
            *    52789 52888: gap of      100 bp
            *    52889    68506: contig of 15618 bp in length
            *    68507 68606: gap of      100 bp
            *    68607   105368: contig of 36762 bp in length
            *   105369 105468: gap of      100 bp
            *   105469   133826: contig of 28358 bp in length
            *   133827 133926: gap of      100 bp
            *   133927   170402: contig of 36476 bp in length
            *   170403 170502: gap of      100 bp
            *   170503   184444: contig of 13942 bp in length.
FEATURES             Location/Qualifiers
     source          1. .184444
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-868P18"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .24646
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    24747. .28060
                     /note="assembly_fragment"
     misc_feature    28161. .33796
                     /note="assembly_fragment"
     misc_feature    33897. .41745
                     /note="assembly_fragment"
     misc_feature    41846. .52788
                     /note="assembly_fragment"
```

```
     misc_feature    52889. .68506
                     /note="assembly_fragment"
     misc_feature    68607. .105368
                     /note="assembly_fragment"
     misc_feature    105469. .133826
                     /note="assembly_fragment"
     misc_feature    133927. .170402
                     /note="assembly_fragment"
     misc_feature    170503. .184444
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    52113 a  41006 c  41428 g  48985 t    912 others
ORIGIN

  Query Match             28.4%;  Score 262.4;  DB 2;  Length 184444;
  Best Local Similarity   99.6%;  Pred. No. 3.2e-54;
  Matches  263;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     621 AAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAATGAGGAGCCACGCC 680
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10772 ACAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAATGAGGAGCCACGCC 10713

Qy     681 TTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTGCAGGCACAGGGTGCCCC 740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10712 TTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTGCAGGCACAGGGTGCCCC 10653

Qy     741 TGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTGTTGCGGGGGCAACATGAGAG 800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10652 TGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTGTTGCGGGGGCAACATGAGAG 10593

Qy     801 CCCCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGTGAAGAGCCCCACCTCGGGGT 860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10592 CCCCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGTGAAGAGCCCCACCTCGGGGT 10533

Qy     861 CACAATAAAGTTGCCTGGTCAGGA 884
           ||||||||||||||||||||||||
Db   10532 CACAATAAAGTTGCCTGGTCAGGA 10509


RESULT 8
HS833B7/c
LOCUS       HS833B7                 86574 bp    DNA     linear   PRI 12-DEC-1999
DEFINITION  Human DNA sequence from clone CTA-833B7 on chromosome 22q12.3-13.2
            Contains the NCF4 gene for cytosolic neutrophil factor 4 (40kD),
            the 5' part of the CSF2RB gene for granulocyte-macrophage
            low-affinity colony stimulating factor 2 receptor beta, ESTs, STSs
            and GSSs, complete sequence.
ACCESSION   AL008637
VERSION     AL008637.1  GI:3136000
KEYWORDS    HTG; colony stimulating factor; CSF2RB; cytosolic neutrophil
            factor; NCF4.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 86574)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 18, 1998 this sequence version replaced gi:2578146.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 22, constructed by the Sanger Centre Chromosome 22
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr22
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep CTA-833B7 is
            from the human BAC library described in U-J. Kim et al. (1996)
            Genomics 34, 213-218.
            VECTOR: pBeloBAC11
            IMPORTANT: This sequence is not the entire insert of clone
            CTA-833B7 It may be shorter because we only sequence overlapping
            sections once, or longer because we arrange for a small overlap
            between neighbouring submissions.
            The true right end of clone CITF22-24E5 is at 100 in this sequence.
            The start of this sequence overlaps with sequence Z82185.
FEATURES             Location/Qualifiers
     source          1. .86574
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="q12.3-13.2"
                     /clone="CTA-833B7"
                     /clone_lib="CIT978SK-A2"
     repeat_region   1. .176
                     /note="AluSp repeat: matches 138. .313 of consensus"
     repeat_region   862. .903
                     /note="MIR repeat: matches 212. .254 of consensus"
     repeat_region   1791. .2095
                     /note="AluJb repeat: matches 1. .307 of consensus"
     misc_feature    complement(2600. .3449)
                     /note="match: GSS: Em:AQ749444"
     repeat_region   3359. .3522
                     /note="MER58A repeat: matches 43. .207 of consensus"
     repeat_region   3532. .3838
                     /note="AluSx repeat: matches 1. .308 of consensus"
     repeat_region   4030. .4288
                     /note="MLT1J repeat: matches 1. .261 of consensus"
     misc_feature    complement(4402. .4679)
                     /note="match: GSS: Em:AQ321484"
     repeat_region   5050. .5195
                     /note="LTR41 repeat: matches 4. .144 of consensus"
     repeat_region   5291. .5370
                     /note="L1ME1 repeat: matches 5526. .5607 of consensus"
     repeat_region   5371. .5676
                     /note="AluSx repeat: matches 1. .310 of consensus"
     repeat_region   5677. .6252
                     /note="L1ME1 repeat: matches 5607. .6163 of consensus"
     repeat_region   6302. .6485
                     /note="L2 repeat: matches 2078. .2265 of consensus"
     repeat_region   6873. .7196
                     /note="54 copies 6 mer gtgtgt 66 conserved"
     repeat_region   6874. .7197
                     /note="27 copies 12 mer 66 conserved"
     repeat_region   6879. .7190
                     /note="13 copies 24 mer 67 conserved"
     repeat_region   6884. .7195
                     /note="156 copies 2 mer tg 67 conserved"
     repeat_region   6886. .7193
                     /note="22 copies 14 mer 67 conserved"
     repeat_region   6907. .7194
                     /note="9 copies 32 mer 69 conserved"
     repeat_region   7203. .7332
                     /note="L2 repeat: matches 2602. .2750 of consensus"
     repeat_region   7347. .7532
                     /note="MIR repeat: matches 2. .191 of consensus"
     repeat_region   7537. .7640
                     /note="MIR repeat: matches 158. .262 of consensus"
     repeat_region   8152. .8228
```

```
                     /note="MIR repeat: matches 184. .260 of consensus"
repeat_region        8253. .8341
                     /note="L2 repeat: matches 2651. .2748 of consensus"
repeat_region        8374. .8479
                     /note="MIR repeat: matches 48. .167 of consensus"
repeat_region        11386. .11858
                     /note="L2 repeat: matches 2164. .2749 of consensus"
gene                 11913. .29288
                     /gene="NCF4"
mRNA                 join(11913. .12573,15415. .15499,16289. .16442,18762. .18832,
                     21785. .21912,23022. .23079,23697. .23795,27023. .27153,
                     27399. .27464,28998. .29288)
                     /gene="NCF4"
                     /product="bK833B7.1 (neutrophil cytosolic factor 4
                     (40kD))"
                     /note="match: cDNAs: Em:AB002665 Em:X77094 Em:U59488;
                     match: ESTs: Em:AA177839 Em:AI007048 Em:AA465462
                     Em:AA465389 Em:AA485518 Em:AI381940 Em:AA744805
                     Em:AA975113 Em:AA969460 Em:AA948430 Em:AI435296
                     Em:AI299103 Em:AA648472 Em:AA702857 Em:AI088359 Em:D20144
                     Em:AA909156 Em:AA688071 Em:W95229 Em:AI439568"
                     /evidence=not_experimental
repeat_region        11982. .12109
                     /note="L2 repeat: matches 2356. .2502 of consensus"
misc_feature         join(12483. .12573,15415. .15499,16289. .16442,18762. .18832,
                     21785. .21912,23022. .23079,23697. .23795,27023. .27464,
                     28998. .29288)
                     /gene="NCF4"
                     /note="match: STS: Em:X77094"
CDS                  join(12542. .12573,15415. .15499,16289. .16442,18762. .18832,
                     21785. .21912,23022. .23079,23697. .23795,27023. .27153,
                     27399. .27464,28998. .29193)
                     /gene="NCF4"
                     /note="match: proteins: Tr:O60808 Tr:P97369"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="bK833B7.1 (neutrophil cytosolic factor 4
                     (40kD))"
                     /protein_id="CAA15486.1"
                     /db_xref="GI:3136001"
                     /db_xref="SPTREMBL:O60808"
                     /translation="MAVAQQLRAESDFEQLPDDVAISANIADIEEKRGFTSHFVFVIE
                     VKTKGGSKYLIYRRYRQFHALQSKLEERFGPDSKSSALACTLPTLPAKVYVGVKQEIA
                     EMRIPALNAYMKSLLSLPVWVLMDEDVRIFFYQSPYDSEQVPQALRRLRPRTRKVKSV
                     SPQGNSVDRMAAPRAEALFDFTGNSKLELNFKAGDVIFLLSRINKDWLEGTVRGATGI
                     FPLSFVKILKDFPEEDDPTNWLRCYYYEDTISTIKDIAVEEDLSSTPLLKDLLELTRR
                     EFQREDIALNYRDAEGDLVRLLSDEDVALMVRQARGLPSQKRLFPWKLHITQKDNYRV
                     YNTMP"
repeat_region        12766. .12882
                     /note="MIR repeat: matches 20. .136 of consensus"
repeat_region        12944. .13247
                     /note="L2 repeat: matches 2425. .2750 of consensus"
repeat_region        13315. .13403
                     /note="MIR repeat: matches 53. .143 of consensus"
repeat_region        14202. .14242
                     /note="L2 repeat: matches 2701. .2741 of consensus"
repeat_region        14289. .14336
                     /note="2 copies 24 mer 95 conserved"
repeat_region        14545. .14625
                     /note="L2 repeat: matches 2637. .2728 of consensus"
repeat_region        14656. .14743
                     /note="MIR repeat: matches 89. .179 of consensus"
repeat_region        14842. .14977
                     /note="L2 repeat: matches 2213. .2350 of consensus"
repeat_region        14983. .15050
                     /note="L2 repeat: matches 2638. .2710 of consensus"
repeat_region        15130. .15225
                     /note="3 copies 32 mer 95 conserved"
repeat_region        15616. .15815
                     /note="MER20 repeat: matches 1. .218 of consensus"
repeat_region        16065. .16102
                     /note="19 copies 2 mer tt 81 conserved"
repeat_region        16171. .16218
                     /note="16 copies 3 mer cct 77 conserved"
repeat_region        16173. .16214
                     /note="7 copies 6 mer tcctcc 81 conserved"
repeat_region        17561. .17987
                     /note="L2 repeat: matches 1251. .1706 of consensus"
repeat_region        18193. .18242
                     /note="L2 repeat: matches 2702. .2750 of consensus"
repeat_region        18243. .18412
                     /note="MER5B repeat: matches 1. .175 of consensus"
repeat_region        18413. .18473
                     /note="L2 repeat: matches 2638. .2702 of consensus"
misc_feature         complement(18442. .18745)
                     /note="match: STS: Em:HSA046ZA5; match: STS: Em:Z51609"
repeat_region        18481. .18516
                     /note="3 copies 12 mer 94 conserved"
misc_feature         complement(18698. .19137)
                     /note="match: GSS: Em:AQ499165"
repeat_region        19252. .19406
                     /note="MER91A repeat: matches 11. .193 of consensus"
repeat_region        19642. .19730
                     /note="MIR repeat: matches 57. .145 of consensus"
repeat_region        20478. .20578
                     /note="MIR repeat: matches 9. .104 of consensus"
```

```
 Query Match             18.6%;  Score 171.6;  DB 9;  Length 86574;
 Best Local Similarity   71.9%;  Pred. No. 9.8e-32;
 Matches  299;  Conservative    0;  Mismatches   74;  Indels   43;  Gaps    4;

Qy     428 AAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATATATCTC 487
           | ||  | | |||||| ||||| |||||||||||||||||| ||||||||||    ||||
Db   16224 AGGGAAAAGGGGAGGAGGAGGAGGAAGATGATGAGGATAAGGATGCAGAAGGGTCTC 16165

Qy     488 TGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCG 547
           ||||| || | | ||||||    |||||||||||||||| |||||||||||| |||||
Db   16164 TGGAGAAGGAGACCCCTGT----CAAGTCAAAAGGCTGGCGCCCCAGAAGCAGACGAGCT 16109

Qy     548 TGGCTAAGAAAAAAAAGCTGGAAAAAGAAGA--------------AGAGGAAATAAGAGC 593
           | || || |||||| |   | ||| | || |              ||||||| | ||| |
Db   16108 TTGCGAAAAAAAAAGAAAAGAAAAGAAAAAAAAAACTGGAAAAAGAGAGGAAGTGAGACC 16049

Qy     594 CAGCGTTAGAGACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCC 653
            || ||||||| |||||||||  |||| ||||||||  ||||  |||  |
Db   16048 TAGTGTTAGAGGCAAGAGCCCTTGGAAATAGGCCAAACACACATTCAGGCC--------- 15998

Qy     654 AGGATTCAAGAAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTG 713
                 |||||||||||||||| ||||||| || || |||||| |||| ||| |||
Db   15997 ------CAAGAAATGAGGAGCCATGCCTTGGAGGTCATGGTGCAGAGTGAACCTGCCCCT 15944

Qy     714 GGCTGTGCTGCAGGCACAGGGTGCCCCTGTCCAGCCCCTCCACCTGTGTCTGAATGCAAC 773
           ||||||||||  ||  |||| |||          |||||||| ||||||||||   ||
Db   15943 GGCTGTGCTGTGGGTGCAGGATGC----------CCCCTCCATCTGTGTCTGAATATGAC 15894

Qy     774 AGGGGTGTTGCGGGGGCAACATGAGAGCCCCTCACCCCCAACTCTCCACTTTCAGG 829
           |||||||||| |||||||||  | || |||||||||| ||||||||  |  |||
Db   15893 AGGGGTGTTGTGGGGGCAACACTAAAGTCCCTCACCCCCGACTCTCCAGGTGGAGG 15838
```

```
RESULT 9
AC108486
LOCUS       AC108486               135206 bp    DNA     linear   HTG 29-JAN-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-114C6, WORKING DRAFT SEQUENCE,
            10 unordered pieces.
ACCESSION   AC108486
VERSION     AC108486.1  GI:18390285
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 135206)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
```

```
REFERENCE   2  (bases 1 to 135206)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     -------------- Genome Center
              Center: University of Washington Genome Center
              Center Code: UWGC
              Web site: http://www.genome.washington.edu
              Contact: uwgchtgs@u.washington.edu
            -------------- Project Information
              Center project name: chr-3
              Center clone name: RP11-114C6 (bc0623)
            ------------- Summary Statistics
              Sequencing vector: plasmid; 100% of reads
              Chemistry: Dye-terminator ET; 20% of reads
              Chemistry: Dye-terminator Big Dye; 80% of reads
              Assembly program: Phrap; version 0.990319
              Consensus quality: 131485 bases at least Q40
              Consensus quality: 133037 bases at least Q30
              Consensus quality: 133846 bases at least Q20
              Insert size: 180200; 27.0% error; agarose-fp
              Insert size: 134306; sum-of-contigs
              Quality coverage: 7.6x in Q20 bases; agarose-fp
              Quality coverage: 10.2x in Q20 bases; sum-of-contigs
            -------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2717: contig of 2717 bp in length
            *     2718     2817: gap of unknown length
            *     2818    11010: contig of 8193 bp in length
            *    11011    11110: gap of unknown length
            *    11111    18451: contig of 7341 bp in length
            *    18452    18551: gap of unknown length
            *    18552    28925: contig of 10374 bp in length
            *    28926    29025: gap of unknown length
            *    29026    39061: contig of 10036 bp in length
            *    39062    39161: gap of unknown length
            *    39162    46549: contig of 7388 bp in length
            *    46550    46649: gap of unknown length
            *    46650    56395: contig of 9746 bp in length
            *    56396    56495: gap of unknown length
            *    56496    79127: contig of 22632 bp in length
            *    79128    79227: gap of unknown length
            *    79228    99937: contig of 20710 bp in length
            *    99938   100037: gap of unknown length
            *   100038   135206: contig of 35169 bp in length.
FEATURES             Location/Qualifiers
     source          1. .135206
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-114C6"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1. .2717
                     /note="assembly_name:Contig14"
     misc_feature    2818. .11010
                     /note="assembly_name:Contig15"
     misc_feature    11111. .18451
                     /note="assembly_name:Contig16"
     misc_feature    18552. .28925
                     /note="assembly_name:Contig17"
     misc_feature    29026. .39061
                     /note="assembly_name:Contig18"
     misc_feature    39162. .46549
                     /note="assembly_name:Contig19"
     misc_feature    46650. .56395
                     /note="assembly_name:Contig20"
     misc_feature    56496. .79127
                     /note="assembly_name:Contig21"
     misc_feature    79228. .99937
                     /note="assembly_name:Contig22"
     misc_feature    100038. .135206
                     /note="assembly_name:Contig23"
BASE COUNT    43564 a  23724 c  24336 g  42625 t    957 others
ORIGIN

  Query Match             13.2%;  Score 122;  DB 2;  Length 135206;
  Best Local Similarity   61.7%;  Pred. No. 2e-19;
  Matches  369;  Conservative    0;  Mismatches  155;  Indels   74;  Gaps    8;

Qy     317 TTTCTCCCCCAGTTACTTTCCAGCTCCGGGCTGGCTCAGGACCCGTGTTCCTCAGTGGCC 376
           ||||||| ||| | |||||  | ||| | ||| ||||   | || ||||| |||||| |
Db   22428 TTTCTCCTCCAATCACTTTTTAACTCTGAGCT-GCTCCAAATCCCTGTTCTTCAGTGACT 22486

Qy     377 AGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAGGAAGAAGAAGGGGAGG 436
           |   | |||||||| | | || |||||||||||| ||||| ||  | || | || ||| |
Db   22487 AAAGATGTTATGAAACTTTAGCCCTAACCTGGGAAGAGGAAGAAAAGGAGGGAGAGGAAG 22546

Qy     437 AGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATATATCTCTGGAGGAGC 496
           | ||||| |||| |||                      | |  | |||||  |||||
Db   22547 ACGAGGAGGAGGGAGA----------------------ATAGTTGTCTCTAAAGGAGG 22582

Qy     497 AAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCGTGGCTAAGA 556
           ||| | || ||||||| |||||| |   ||  ||| |||||| | | | |   |||||||
Db   22583 AAACCTCTATCAAACATGTCAAAGGTGGGGGACCCAAGAAGCTGAC-AACACTGCTAAGA 22641

Qy     557 AAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTAGAGACAAGAGCCCTG 616
             ||  |||||||||||          ||      |    ||| ||||||||||||||
Db   22642 GCAAGGAGCTGGAAAAAG----------GATTGAGATCCATCTTAAAGACAAGAGCCCTG 22691

Qy     617 TGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAATGAGGAGCCA 676
           ||| ||||  |||| ||| |  || | | ||||| ||||| | |||||||| ||||||||
Db   22692 TGAGAAAGAACAAACCCATACTCAAACCTAAGAAACCAGGTTCCAAGAAATAAGGAGCCA 22751

Qy     677 ------------------------------CGCCTTGGGGGGCACGGTGCAAAGTGGGCC 706
                                           ||||  | |  ||   |||| ||||||||
Db   22752 GGAACTGCCTGACTGCGGTGATCAAGAGAGGGCCTGTGAGACCAGCATGCAGAGTGGGCC 22811

Qy     707 TTCCCTGGGCTGTGCTGCAGGCACAGGGTGCCCCTGTCC---AGCCCCTCCACCTGTGTC 763
           |||| ||||||||||||  | ||  |||||  ||  | |     ||| |||| ||| |||
Db   22812 TTCCTTGGGCTGTGCTGTGGTCATGGGGTGTACCCCTACCTTGCCCCTTCCATCTGAGTC 22871

Qy     764 TGAATGCAACAGGGGTGTTGCGGGGGCAACATGAGA----GCCCCTCACCCCCAACTCTC 819
           | |||  ||  || |||||  || |  |     | |     | | ||  | | ||||||||
Db   22872 TAAATAGAATGGGAGTGTTTTGGAGAAAGTGACAAAAGATCCTCTTCTCACTCAACTCTC 22931

Qy     820 CACTTTCAGGAGGCCCCCA-GTGAAGAGCCCCACCTCGGGGTCACAATAAAGTTGCCT 876
           |  | | |||||||||    | ||||||||  ||| | |||  ||||||||||  ||
Db   22932 CCATGTGAGGAGGCCCTGGTGAGAAGAGCCTGACCACAGGGAAACAATAAAGTTTGCT 22989


RESULT 10
AC093557
LOCUS       AC093557               149940 bp    DNA     linear   PRI 24-MAR-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-56P22, complete sequence.
ACCESSION   AC093557 AC022225
VERSION     AC093557.3  GI:19698701
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 149940)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 149940)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J., Ivey,R.G. and
```

```
            Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 149940)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Mar 24, 2002 this sequence version replaced gi:18087686.
            ------------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
                Drafting Center: BCM
            ------------- Project Information
                Center project name: chr-3
                Center clone name: RP11-56P22 (bc0176)
            ------------ Summary Statistics
                Sequencing vector: M13; L08821; 52% of reads
                Sequencing vector: plasmid; 48% of reads
                Sequencing vector: plasmid; L08752; 0% of reads
                Chemistry: Dye-terminator ET; 17% of reads
                Chemistry: Dye-primer Bodipy; 1% of reads
                Chemistry: Dye-terminator Big Dye; 82% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 149854 bases at least Q40
                Consensus quality: 149935 bases at least Q30
                Consensus quality: 149940 bases at least Q20
                Insert size: 149940; sum-of-contigs
                Quality coverage: 8.3x in Q20 bases; sum-of-contigs
            -------------.
            Overlapping Sequences:
                5':  RP11-237P23 (UWGC:bc0783) AC112223
                3':  RP11-114C6 (UWGC:bc0623) AC108486
            --------------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.

            This sequence was finished as follows unless otherwise noted:
            all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., Phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            --------------------
            Sequence Validation:
             This sequence has been validated by Multiple Complete Digest
             fingerprinting. Comparison of the experimentally derived digest
             fragments with sequence-predicted fragments is given below.
             The electronically-digested sequence consists of both insert and
             vector, in order to accurately represent the entire circular BAC.
             Small fragments below a variable cutoff (approximately 400-800 bp)
             are not resolved in the fingerprint and hence do not appear
             in the table. There are no significant remaining discrepancies
             between the experimental and predicted values. Uniquely ordered
             fragments are separated by dashed lines.
                  EcoRI                HindIII              BglII
            -------------------  -------------------  -------------------
            SeqDerMap  FngrPrnt  SeqDerMap  FngrPrnt  SeqDerMap  FngrPrnt
            ---------  --------  ---------  --------  ---------  --------

            ---------            ---------            ---------
```

8696	8608	3232	3238	10777	10725
6	<800	6382	6740	2067	2070
1938	1915	512	<800	8879	9137
2860	2886	449	<800	4717	4762
665	<800	3551	3535	1161	1132
1137	1166	66	<800	3316	3323
3903	3847	4029	4016	7117	7092
357	<800	1886	1892	601	<800
2519	2497	379	<800	14637	14477
212	<800	5099	5054	23604	23653
3478	3509	2949	3028	2916	2926
2469	2497	5004	5054	2176	2167
873	900	1085	1067	556	<800
279	<800	1671	1632	3689	3724
5376	5312	393	<800	2880	2926
2045	2067	8291	8472	7884	7435
1963	1915	1962	1892	565	<800
1172	1166	7255	7042	3478	3460
1272	1307	2118	2170	2658	2665
1083	1037	1169	1168	741	744
9619	9736	1330	1292	2895	2926
1642	1625	235	<800	11488	11677
199	<800	1294	1292	1498	1521
3489	3509	857	901	2400	2416
4629	4644	2838	2805	5145	5305
8423	8608	4757	4713	1202	1162
3204	3186	2331	2337	30	<800
6821	6882	1896	1892	1839	1807
1033	1037	761	768	1978	1930
4701	4644	1612	1632	2469	2416
17125	16431	2232	2170	2637	2665
2527	2497	6677	6740	808	830
2974	2886	201	<800	6888	6667
3166	3186	677	<800	4739	4762
4025	4005	265	<800	1346	1323
14230	14884	3833	3824	2429	2416
5750	5785	7946	7558	12	<800

279	<800	2301	2337	655	<800
1323	1307	2753	2805	1341	1323
12	<800	490	<800	1563	1521
2101	2067	177	<800	690	<800
4290	4236	6836	6740	998	1005
1154	1166	6767	6740	509	<800
766	801	5595	5684		
6886	6882	5959	5945		
7307	7318	10600	10609		
		5616	5684		
		5062	5054		
		111	<800		
		4421	4433		
		1890	1892		
		4176	4156		

```
FEATURES             Location/Qualifiers
     source          1. .149940
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-56P22"
                     /clone_lib="RPCI human BAC library 11"

  Query Match             13.0%;  Score 120.4;  DB 9;  Length 149940;
  Best Local Similarity   61.5%;  Pred. No. 5.1e-19;
  Matches  368;  Conservative    0;  Mismatches  156;  Indels   74;  Gaps    8;

Qy     317 TTTCTCCCCCAGTTACTTTCCAGCTCCGGGCTGGCTCAGGACCCGTGTTCCTCAGTGGCC 376
           ||||||| ||| | |||||  | ||| | ||| ||||   | || ||||| |||||| |
Db   79005 TTTCTCCTCCAATCACTTTTTAACTCTGAGCT-GCTCCAAATCCCTGTTCTTCAGTGACT 79063

Qy     377 AGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAGGAAGAAGAAGGGGAGG 436
           |   | |||||||| | | || |||||||||||||| ||||| ||  | || | || ||| |
Db   79064 AAAGATGTTATGAAACTTTAGCCCTAACCTGGGAAGAGGAAGAAAAGGAGGGAGAGGAAG 79123

Qy     437 AGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATATATCTCTGGAGGAGC 496
           | ||||| |||| |||                      | |  | |||||  |||||
Db   79124 ACGAGGAGGAGGGAGA----------------------ATAGTTGTCTCTAAAGGAGG 79159

Qy     497 AAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCGTGGCTAAGA 556
           ||| | || ||||||| |||||| |   ||  ||| |||||| | | | |   |||||||
Db   79160 AAACCTCTATCAAACATGTCAAAGGTGGGGGACCCAAGAAGCTGAC-AACACTGCTAAGA 79218

Qy     557 AAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTAGAGACAAGAGCCCTG 616
           |  ||||||||||||          ||      |    ||| ||||||||||||||||
Db   79219 GCGAGGAGCTGGAAAAAG----------GATTGAGATCCATCTTAAAGACAAGAGCCCTG 79268

Qy     617 TGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAATGAGGAGCCA 676
           ||| ||||  |||| ||| |  || | | ||||| ||||| | |||||||| ||||||||
Db   79269 TGAGAAAGAACAAACCCATACTCAAACCTAAGAAACCAGGTTCCAAGAAATAAGGAGCCA 79328

Qy     677 ------------------------------CGCCTTGGGGGGCACGGTGCAAAGTGGGCC 706
                                        ||||  | |  ||   |||| ||||||||
Db   79329 GGAACTGCCTGACTGCGGTGATCAAGAGAGGGCCTGTGAGACCAGCATGCAGAGTGGGCC 79388

Qy     707 TTCCCTGGGCTGTGCTGCAGGCACAGGGTGCCCCTGTCC---AGCCCCTCCACCTGTGTC 763
           |||| ||||||||||||  | ||  |||||  ||  | |     ||| |||| ||| |||
```

```
Db   79389 TTCCTTGGGCTGTGCTGTGGTCATGGGGTGTACCCCTACCTTGCCCCTTCCATCTGAGTC 79448

Qy     764 TGAATGCAACAGGGGTGTTGCGGGGGCAACATGAGA----GCCCCTCACCCCCAACTCTC 819
           | |||  ||  || |||||  || |  |     | |     | | || | | ||||||||
Db   79449 TAAATAGAATGGGAGTGTTTTGGAGAAAGTGACAAAAGATCCTCTTCTCACTCAACTCTC 79508

Qy     820 CACTTTCAGGAGGCCCCCA-GTGAAGAGCCCCACCTCGGGGTCACAATAAAGTTGCCT 876
           |  | | |||||||||    | ||||||||  ||| | |||  |||||||||||  ||
Db   79509 CCATGTGAGGAGGCCCTGGTGAGAAGAGCCTGACCACAGGGAAACAATAAAGTTTGCT 79566
```

```
RESULT 11
XLNUPLR
LOCUS       XLNUPLR                  646 bp    mRNA    linear   VRT 10-FEB-1999
DEFINITION  Xenopus laevis mRNA fragment for nucleoplasmin.
ACCESSION   Y00204
VERSION     Y00204.1  GI:64938
KEYWORDS    nucleoplasmin.
SOURCE      Xenopus laevis.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 646)
  AUTHORS   De Robertis,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-1987) de Robertis E., University of California,
            Dept. of Biological Chemistry, University of California, Los
            Angeles, CA 90024, USA
REFERENCE   2  (bases 1 to 646)
  AUTHORS   Burglin,T.R., Mattaj,I.W., Newmeyer,D.D., Zeller,R. and De
            Robertis,E.M.
  TITLE     Cloning of nucleoplasmin from Xenopus laevis oocytes and analysis
            of its developmental expression
  JOURNAL   Genes Dev. 1 (1), 97-107 (1987)
  MEDLINE   88112783
   PUBMED   3428591
FEATURES             Location/Qualifiers
     source          1. .646
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="N7.2"
                     /clone_lib="lambda gt11 cDNA library"
     CDS             <1. .575
                     /codon_start=3
                     /product="nucleoplasmin"
                     /protein_id="CAA68363.1"
                     /db_xref="GI:833629"
                     /db_xref="SWISS-PROT:P05221"
                     /translation="NTSKVEKPVSLIWGCELNEQNKTFAFKIEDEEEKCEHQLALRTV
                     CLGDKAKDEFHIVEIVTQEEGKEKPVPIASLKPSILPMATMVGIELTPPVTFRLKAGS
                     GPVYISGQHVAMEEDYSWAEEEDEGEEEEEEEEDPESPPKAVKRPAATKKAGQAKKKK
                     LDKEDESSEEDSPTKKGKGAGRGRKPAAKK"
     polyA_signal    634. .639
BASE COUNT      222 a    121 c    174 g    129 t
ORIGIN

  Query Match             10.8%;  Score 99.6;  DB 5;  Length 646;
  Best Local Similarity   57.4%;  Pred. No. 4.8e-14;
  Matches  224;  Conservative    0;  Mismatches  154;  Indels   12;  Gaps    2;

Qy      69 GACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGACTTGGACCTTCAGACCCCA 128
           | || |  | ||||| || || || | | || |||    ||| |   | || | |    |
Db      29 GTCCCTTATTTGGGGGTGTGAACTGAATGAGCAGAACAAGACATTTGCGTTTAAAATAGA 88

Qy     129 GCTGGAGGGGAAGCAGAGCTGCAGGCTGTTG---CTTCATACGATTTGCTTGGGGGAGAA 185
           ||||   |  |  |      | ||||   || |  ||  | ||  | ||||| ||
Db      89 AGACGAGGAAGAAAAATGTGAGCATCAGTTGGCACTGCGCACAGTATGTCTTGGGGACAA 148

Qy     186 AGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCAGCAAACCAGGAGGACAAGAA 245
           || || || ||| | ||    || || ||         ||  |  || |||||    ||
Db     149 GGCAAAGGATGAGTTCCACATAGTAGAAAT---------AGTTACACAAGAGGAGGGAAA 199
```

```
Qy     246 GATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCCCATGGTCTCCATGGT 305
           |    | || ||  | ||||||   ||  || | ||  |  | |||||||   |||||||
Db     200 GGAAAAACCAGTTCCGATTGCCAGTCTAAAGCCTTCTATTTTACCCATGGCAACCATGGT 259

Qy     306 AGGAGTGCAGCTTTCTCCCCCAGTTACTTTCCAGCTCCGGGCTGGCTCAGGACCCGTGTT 365
            ||  |  ||||  | || ||||||||||||| | |    |||||  || || || ||||
Db     260 GGGCATTGAGCTGACCCCTCCAGTTACTTTCCGGTTAAAAGCTGGTTCCGGCCCAGTGTA 319

Qy     366 CCTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAGGAAGA 425
           | ||||||| ||  |         |    |||     | ||||  || ||||| || ||
Db     320 CATCAGTGGTCAACACGTAGCAATGGAGGAAGATTACTCATGGGCAGAAGAGGAAGATGA 379

Qy     426 AGAAGGGGAGGAGGAGGAAGAGGAAGAGGA 455
           || ||  ||||||||||||||||||||| ||
Db     380 AGGAGAAGAGGAGGAGGAAGAGGAAGAAGA 409
```

RESULT 12
XLNUPLR2
LOCUS XLNUPLR2 800 bp mRNA linear VRT 12-SEP-1993
DEFINITION Xenopus laevis mRNA for nucleoplasmin.
ACCESSION X04766
VERSION X04766.1 GI:64939
KEYWORDS nucleoplasmin.
SOURCE Xenopus laevis.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 800)
AUTHORS Dingwall,C., Dilworth,S.M., Black,S.J., Kearsey,S.E., Cox,L.S. and Laskey,R.A.
TITLE Nucleoplasmin cDNA sequence reveals polyglutamic acid tracts and a cluster of sequences homologous to putative nuclear localization signals
JOURNAL EMBO J. 6 (1), 69-74 (1987)
MEDLINE 87218476
PUBMED 2884102
FEATURES Location/Qualifiers
source 1. .800
/organism="Xenopus laevis"
/db_xref="taxon:8355"
CDS 114. .716
/note="nucleoplasmin (AA 1-200)"
/codon_start=1
/protein_id="CAA28460.1"
/db_xref="GI:64940"
/db_xref="SWISS-PROT:P05221"
/translation="MASTVSNTSKLEKPVSLIWGCELNEQDKTFEFKVEDDEEKCEHQ
LALRTVCLGDKAKDEFNIVEIVTQEEGAEKSVPIATLKPSILPMATMVGIELTPPVTF
RLKAGSGPLYISGQHVAMEEDYSWAEEEDEGEAEGEEEEEEEEDQESPPKAVKRPAAT
KKAGQAKKKKLDKEDESSEEDSPTKKGKGAGRGRKPAAKK"
polyA_site 800
/note="polyA site"
BASE COUNT 250 a 153 c 222 g 175 t
ORIGIN

Query Match 10.4%; Score 96.2; DB 5; Length 800;
Best Local Similarity 54.9%; Pred. No. 3.4e-13;
Matches 284; Conservative 0; Mismatches 188; Indels 45; Gaps 3;

```
Qy     153 GCTGTTGCTTCATACGATTTGCTTGGGGGAGAAAGCCAAAGAGGAGATGCATCGCGTGGA 212
           | ||  | | |  ||| | ||  |||||||  || || || || ||| |   |    || ||
Db     245 GTTGGCGTTGCGCACGGTGTGTCTGGGGGACAAGGCAAAGGATGAGTTCAACATTGTAGA 304

Qy     213 GATCCTGCCCCCAGCAAACCAGGAGGACAAGAAGATGCAGCCGGTCACCATTGCCTCACT 272
           |||  |  | | ||   |    | ||| |         |  | ||  | |||||| | ||
Db     305 AATCGTTACACAAGAGGAGGGAGCGGAAA---------AATCTGTTCCAATTGCCACTCT 355

Qy     273 CCAGGCCTCAGTCCTCCCCATGGTCTCCATGGTAGGAGTGCAGCTTTCTCCCCCAGTTAC 332
           || | ||  | || |||||||   | |||||  ||  |  ||||  |||| ||||||||
Db     356 AAAGCCTTCTATTCTACCCATGGCAACTATGGTGGGCATTGAGCTGACTCCTCCAGTTAC 415

Qy     333 TTTCCAGCTCCGGGCTGGCTCAGGACCCGTGTTCCTCAGTGGCCAGGAACGTTATGAAGC 392
           ||||| | |    |||||  || || ||  ||| | |||||||  ||
Db     416 TTTCCGGTTAAAAGCTGGTTCCGGCCCACTGTACATCAGTGGTCA--------------- 460

Qy     393 ATCAGACCTAACCTGGGAGGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGA 452
                 |  |   | |  |||||| ||  |   | |||  || ||||||||  ||||
Db     461 ---------ACACGTAGCGATGGAGGAAGATTACTCATGGGCAGAAGAGGAAGATGAAGG 511

Qy     453 GGAAGATGATGAGGATGAGGATGCAGATATATCTCTGGAGGAGCAAAGCCCTGTCAAACA 512
           ||||| ||| |  || |||||  | |||           ||  |  ||   ||   ||||
Db     512 GGAAGCTGAAGGAGAAGAGGAGGAAGAAGAGGAAGAAGATCAAGAATCTCCACCCAAAGC 571

Qy     513 AGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAGCTGGAAAA 572
            || || ||||  | | |    ||  | ||  ||  ||| ||||| || || || || ||
Db     572 TGTAAAGAGGCCTGCGGCTACCAAAAAAGCAGGCCAGGCAAAGAAGAAGAAACTTGACAA 631

Qy     573 AGAAGAAGAGGAAATAAGAGCCAGCGTTAGAGACAAGAGCCCTGTGAAAAAGGCCAAAGC 632
           ||| || |||             |||      ||  | || ||    ||||||| |||||
Db     632 AGAGGACGAG------------AGCTCCGAGGAAGACAGTCCAACCAAAAAGGGCAAAGG 679

Qy     633 CACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAATGA 669
           | |  ||||  |  |||||  |     |||||  |||
Db     680 AGCCGGAAGAGGAAGAAAGCCGGCTGCTAAGAAGTGA 716
```

RESULT 13
I66494/c
LOCUS I66494 7218 bp DNA linear PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES Location/Qualifiers
source 1. .7218
/organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match 10.1%; Score 93.2; DB 6; Length 7218;
Best Local Similarity 3.5%; Pred. No. 2.2e-12;
Matches 14; Conservative 257; Mismatches 125; Indels 0; Gaps 0;

```
Qy     364 TTCCTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAGGAA 423
           ||  |  : ::  ::::: :  : :::: :  :::   ::   ::::::::::::::::
Db    1441 TTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1382

Qy     424 GAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATATA 483
           :::::::::::::::::::::::::::::::::::: :: ::::: ::::: : ::: : :
Db    1381 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1322

Qy     484 TCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCG 543
                 ::::::: ::::    :  ::: ::: :::::  :: :    ::::: ::: :
Db    1321 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1262

Qy     544 AGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTAGA 603
           :: : ::  :::::::::::::  :::::::::::::::::::::: :::::  :: :  :::
Db    1261 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1202

Qy     604 GACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAG 663
           :: :::::    : :::::::::  ::::  : ::  ::::  ::::::  ::::   :::
Db    1201 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1142

Qy     664 AAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTG 723
           ::: :::::  : :    ::::::: : :: : :::: :::        :::  : :  :
Db    1141 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1082
```

```
Qy    724 CAGGCACAGGGTGCCCCTGTCCAGCCCCTCCACCTG 759
          ::: : :::: :     |      ||||| |||||
Db   1081 RRRRRRRRRRRRRRRATCGCAAGCTCCCTCGACCTG 1046


RESULT 14
AC125180
LOCUS       AC125180               198421 bp    DNA     linear   HTG 20-JUN-2002
DEFINITION  Mus musculus chromosome UNK clone RP23-322K1, WORKING DRAFT
            SEQUENCE, 30 unordered pieces.
ACCESSION   AC125180
VERSION     AC125180.1  GI:21490699
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 198421)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 198421)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT
            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------- Project Information ----------
            Center project name: M_BA0322K01
            ----------
            ---------- Summary Statistics  ----------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 186437 bases at least Q40
            Consensus quality: 189477 bases at least Q30
            Consensus quality: 191697 bases at least Q20
            Insert size: 203000; agarose-fp
            Insert size: 196692; sum-of-contigs
            Quality coverage: 4.53 in Q20 bases; agarose-fp
            Quality coverage: 4.25 in Q20 bases; sum-of-contigs
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 30 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1076: contig of 1076 bp in length
            *     1077     1176: gap of unknown length
            *     1177     2506: contig of 1330 bp in length
            *     2507     2606: gap of unknown length
            *     2607     4336: contig of 1730 bp in length
            *     4337     4436: gap of unknown length
            *     4437     5587: contig of 1151 bp in length
            *     5588     5687: gap of unknown length
            *     5688     6909: contig of 1222 bp in length
            *     6910     7009: gap of unknown length
            *     7010     8416: contig of 1407 bp in length
            *     8417     8516: gap of unknown length
            *     8517    10353: contig of 1837 bp in length
            *    10354    10453: gap of unknown length
            *    10454    12880: contig of 2427 bp in length
            *    12881    12980: gap of unknown length
            *    12981    15704: contig of 2724 bp in length
            *    15705    15804: gap of unknown length
            *    15805    18535: contig of 2731 bp in length
            *    18536    18635: gap of unknown length
            *    18636    21653: contig of 3018 bp in length
            *    21654    21753: gap of unknown length
            *    21754    23575: contig of 1822 bp in length
            *    23576    23675: gap of unknown length
            *    23676    26531: contig of 2856 bp in length
            *    26532    26631: gap of unknown length
            *    26632    30435: contig of 3804 bp in length
            *    30436    30535: gap of unknown length
            *    30536    35803: contig of 5268 bp in length
            *    35804    35903: gap of unknown length
            *    35904    41090: contig of 5187 bp in length
            *    41091    41190: gap of unknown length
            *    41191    46181: contig of 4991 bp in length
            *    46182    46281: gap of unknown length
            *    46282    53388: contig of 7107 bp in length
            *    53389    53488: gap of unknown length
            *    53489    61060: contig of 7572 bp in length
            *    61061    61160: gap of unknown length
            *    61161    68114: contig of 6954 bp in length
            *    68115    68214: gap of unknown length
            *    68215    73516: contig of 5302 bp in length
            *    73517    73616: gap of unknown length
            *    73617    80218: contig of 6602 bp in length
            *    80219    80318: gap of unknown length
            *    80319    87536: contig of 7218 bp in length
            *    87537    87636: gap of unknown length
            *    87637    97222: contig of 9586 bp in length
            *    97223    97322: gap of unknown length
            *    97323   105689: contig of 8367 bp in length
            *   105690   105789: gap of unknown length
            *   105790   123827: contig of 18038 bp in length
            *   123828   123927: gap of unknown length
            *   123928   138453: contig of 14526 bp in length
            *   138454   138553: gap of unknown length
            *   138554   160961: contig of 22408 bp in length
            *   160962   161061: gap of unknown length
            *   161062   197994: contig of 36933 bp in length
            *   197995   198094: gap of unknown length
            *   198095   198421: contig of 327 bp in length.
FEATURES             Location/Qualifiers
     source          1. .198421
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-322K1"
     misc_feature    1. .1076
                     /note="assembly_name:Contig24"
     misc_feature    1177. .2506
                     /note="assembly_name:Contig28"
     misc_feature    2607. .4336
                     /note="assembly_name:Contig29"
     misc_feature    4437. .5587
                     /note="assembly_name:Contig30"
     misc_feature    5688. .6909
                     /note="assembly_name:Contig33"
     misc_feature    7010. .8416
                     /note="assembly_name:Contig34"
     misc_feature    8517. .10353
                     /note="assembly_name:Contig35"
     misc_feature    10454. .12880
                     /note="assembly_name:Contig36"
     misc_feature    12981. .15704
                     /note="assembly_name:Contig37"
     misc_feature    15805. .18535
                     /note="assembly_name:Contig38"
     misc_feature    18636. .21653
                     /note="assembly_name:Contig39"
     misc_feature    21754. .23575
                     /note="assembly_name:Contig40"
```

```
     misc_feature    23676. .26531
                     /note="assembly_name:Contig41"
     misc_feature    26632. .30435
                     /note="assembly_name:Contig42"
     misc_feature    30536. .35803
                     /note="assembly_name:Contig43"
     misc_feature    35904. .41090
                     /note="assembly_name:Contig44"
     misc_feature    41191. .46181
                     /note="assembly_name:Contig45"
     misc_feature    46282. .53388
                     /note="assembly_name:Contig46"
     misc_feature    53489. .61060
                     /note="assembly_name:Contig47"
     misc_feature    61161. .68114
                     /note="assembly_name:Contig48"
     misc_feature    68215. .73516
                     /note="assembly_name:Contig49"
     misc_feature    73617. .80218
                     /note="assembly_name:Contig50"
     misc_feature    80319. .87536
                     /note="assembly_name:Contig51"
     misc_feature    87637. .97222
                     /note="assembly_name:Contig52"
     misc_feature    97323. .105689
                     /note="assembly_name:Contig53"
     misc_feature    105790. .123827
                     /note="assembly_name:Contig54"
     misc_feature    123928. .138453
                     /note="assembly_name:Contig55"
     misc_feature    138554. .160961
                     /note="assembly_name:Contig56"
     misc_feature    161062. .197994
                     /note="assembly_name:Contig57"
     misc_feature    198095. .198421
                     /note="assembly_name:Contig20"
BASE COUNT    55859 a  43408 c  41991 g  54232 t   2931 others
ORIGIN

  Query Match             9.8%;  Score 90.6;  DB 2;  Length 198421;
  Best Local Similarity   79.1%;  Pred. No. 1.3e-11;
  Matches  121;  Conservative    0;  Mismatches   29;  Indels    3;  Gaps    1;

Qy     403 ACCTGGGAGGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGAT 462
           || | |||   |     |||||| || | ||| |||||||||||||| |||||||| |||
Db 111479 ACTTCGGACCTGACCTGGGAAGATGACGAGGAAGAGGAGGAAGAGGAGGAGGAAGAGGAT 111538

Qy     463 GAGGATGAGGATGCAGATATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGG 522
           || ||||||||||||||||||| || ||||||   |  |||||||||||||||||||||
Db 111539 GAAGATGAGGATGCAGATATATCGCTAGAGGAG---ATACCTGTCAAACAAGTCAAAAGG 111595

Qy     523 CTGGTGCCCCAGAAGCAGGCGAGCGTGGCTAAG 555
            |||  ||||||||||||  |||| | || |||
Db 111596 GTGGCTCCCCAGAAGCAGATGAGCATAGCAAAG 111628


RESULT 15
AC122303
LOCUS       AC122303               190604 bp    DNA     linear   HTG 23-MAY-2002
DEFINITION  Mus musculus chromosome UNK clone RP23-279L13, WORKING DRAFT
            SEQUENCE, 17 unordered pieces.
ACCESSION   AC122303
VERSION     AC122303.1  GI:21105162
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 190604)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 190604)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT
            ---------------- Genome Center ---------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information ------------------
            Center project name: M_BA0279L13
            -----------------------------------------------------
            ------------- Summary Statistics  -------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 184891 bases at least Q40
            Consensus quality: 186045 bases at least Q30
            Consensus quality: 186920 bases at least Q20
            Insert size: 193000; agarose-fp
            Insert size: 189444; sum-of-contigs
            Quality coverage: 6.63 in Q20 bases; agarose-fp
            Quality coverage: 6.43 in Q20 bases; sum-of-contigs
            -----------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1661: contig of 1661 bp in length
            *     1662     1761: gap of unknown length
            *     1762     3532: contig of 1771 bp in length
            *     3533     3632: gap of unknown length
            *     3633     5569: contig of 1937 bp in length
            *     5570     5669: gap of unknown length
            *     5670     8287: contig of 2618 bp in length
            *     8288     8387: gap of unknown length
            *     8388     9782: contig of 1395 bp in length
            *     9783     9882: gap of unknown length
            *     9883    11879: contig of 1997 bp in length
            *    11880    11979: gap of unknown length
            *    11980    14172: contig of 2193 bp in length
            *    14173    14272: gap of unknown length
            *    14273    20234: contig of 5962 bp in length
            *    20235    20334: gap of unknown length
            *    20335    28899: contig of 8565 bp in length
            *    28900    28999: gap of unknown length
            *    29000    35723: contig of 6724 bp in length
            *    35724    35823: gap of unknown length
            *    35824    41909: contig of 6086 bp in length
            *    41910    42009: gap of unknown length
            *    42010    50557: contig of 8548 bp in length
            *    50558    50657: gap of unknown length
            *    50658    61265: contig of 10608 bp in length
            *    61266    61365: gap of unknown length
            *    61366    68507: contig of 7142 bp in length
            *    68508    68607: gap of unknown length
            *    68608    84968: contig of 16361 bp in length
            *    84969    85068: gap of unknown length
            *    85069   116294: contig of 31226 bp in length
            *   116295   116394: gap of unknown length
            *   116395   190604: contig of 74210 bp in length.
FEATURES             Location/Qualifiers
     source          1. .190604
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
```

```
                     /clone="RP23-279L13"
  misc_feature       1. .1661
                     /note="assembly_name:Contig20"
  misc_feature       1762. .3532
                     /note="assembly_name:Contig22
                     clone_end:SP6
                     vector_side:right"
  misc_feature       3633. .5569
                     /note="assembly_name:Contig23"
  misc_feature       5670. .8287
                     /note="assembly_name:Contig24"
  misc_feature       8388. .9782
                     /note="assembly_name:Contig25"
  misc_feature       9883. .11879
                     /note="assembly_name:Contig26"
  misc_feature       11980. .14172
                     /note="assembly_name:Contig27"
  misc_feature       14273. .20234
                     /note="assembly_name:Contig28"
  misc_feature       20335. .28899
                     /note="assembly_name:Contig29"
  misc_feature       29000. .35723
                     /note="assembly_name:Contig30"
  misc_feature       35824. .41909
                     /note="assembly_name:Contig31"
  misc_feature       42010. .50557
                     /note="assembly_name:Contig32"
  misc_feature       50658. .61265
                     /note="assembly_name:Contig33"
  misc_feature       61366. .68507
                     /note="assembly_name:Contig34"
  misc_feature       68608. .84968
                     /note="assembly_name:Contig35
                     clone_end:T7
                     vector_side:left"
  misc_feature       85069. .116294
                     /note="assembly_name:Contig36"
  misc_feature       116395. .190604
                     /note="assembly_name:Contig37"
BASE COUNT    53210 a  41567 c  41640 g  52583 t   1604 others
ORIGIN

  Query Match             9.0%;  Score 83.2;  DB 2;  Length 190604;
  Best Local Similarity   61.6%;  Pred. No. 8.6e-10;
  Matches  133;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy     377 AGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAGGAAGAAGAAGGGGAGG 436
           | |||    | |||| |  |||   |     ||||||||||||||| || || | |||||
Db   49745 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 49804

Qy     437 AGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATATATCTCTGGAGGAGC 496
           ||||||| ||||| ||||| || || ||||| ||||| |  ||   |     |||||||
Db   49805 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAAGAGGAGGAGGAGG 49864

Qy     497 AAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCGTGGCTAAGA 556
           || |    |    | |||  | ||||  |  |    || || ||| |   | ||   |
Db   49865 AAGGAGGAGAAGGAGAAGGAAGAAGGAAGAAGAAGGAGGAGGAGGAGGAGGAAGAAG 49924

Qy     557 AAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAG 592
           || || |    ||| ||||||||||||||  |  ||
Db   49925 AAGAAGAAGAAGAAGAAGAAGAAGAGGAAGAAGAAG 49960


Search completed: January 24, 2003, 18:22:30
Job time : 46614 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 23, 2003, 14:00:24 ; Search time 310 Seconds
(without alignments)
6712.408 Million cell updates/sec

Title: US-09-844-864-16
Perfect score: 924
Sequence: 1 cagcccgcttctctgcccgg.....tttgcggccgcaagcttatg 924

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	858	92.9	1048	21	AAA72432	Human nucleic acid
	2	342.8	37.1	1019	21	AAD00296	Mouse oocyte-speci
	3	337	36.5	423	23	AAD23440	Human lung tumour-
c	4	171.6	18.6	86574	24	ABK83560	Human cDNA differe
	5	131.6	14.2	1481	21	AAD00297	Mouse ovary-specif
	6	72.4	7.8	416	22	AAL01804	Human reproductive
	7	72.4	7.8	416	23	ABL97097	Human testicular a
	8	70.2	7.6	49999	20	AAZ23891	Murine LOBO genomi
	9	70.2	7.6	49999	20	AAZ23896	Murine LOBO homolo
	10	69.2	7.5	33923	22	AAK67071	Human immune/haema
	11	68.8	7.4	341	23	AAS90676	DNA encoding novel
	12	67.8	7.3	512	24	ABQ54697	Human ovarian anti
	13	66.4	7.2	291	23	AAS68926	DNA encoding novel
	14	65.8	7.1	1680	23	AAS68538	DNA encoding novel
c	15	65.2	7.1	8165	22	AAK82672	Human immune/haema
	16	64	6.9	510	23	AAS69539	DNA encoding novel
	17	64	6.9	510	23	AAS71141	DNA encoding novel
	18	64	6.9	510	23	AAS90687	DNA encoding novel
	19	63.8	6.9	591	23	AAS70521	DNA encoding novel
	20	63.8	6.9	591	23	AAS70706	DNA encoding novel
	21	63.8	6.9	591	23	AAS90721	DNA encoding novel
	22	63.8	6.9	1416	23	AAS67163	DNA encoding novel
	23	63.4	6.9	575	22	ABA50472	Human breast cell
	24	63.4	6.9	575	22	ABA68422	Human foetal liver
	25	63.4	6.9	575	22	ABA35413	Probe #13879 for g
	26	63.4	6.9	575	22	AAK16793	Human brain expres
	27	63.4	6.9	575	22	AAK42567	Human bone marrow
	28	63.4	6.9	575	22	AAI23315	Probe #13248 for g
	29	63.4	6.9	575	22	AAI48636	Probe #17322 used
	30	63.4	6.9	575	22	AAI08956	Probe #8947 used t
	31	63.4	6.9	575	24	ABS16615	Human genome-deriv
	32	63.4	6.9	1969	22	ABA45341	Human breast cell
	33	63.4	6.9	1969	22	ABA55830	Human foetal liver
	34	63.4	6.9	1969	22	ABA25506	Probe #3972 for ge
	35	63.4	6.9	1969	22	AAK04048	Human brain expres
	36	63.4	6.9	1969	22	AAK29533	Human bone marrow
	37	63.4	6.9	1969	22	AAI14105	Probe #4038 for ge
	38	63.4	6.9	1969	22	AAI35486	Probe #4172 used t
	39	63.4	6.9	1969	22	AAI03958	Probe #3949 used t
	40	63.4	6.9	1969	24	ABS04084	Human genome-deriv
c	41	63.4	6.9	51259	18	AAX83007	Partial mouse WRN
	42	63.2	6.8	14798	24	ABL33033	Human immune syste
	43	63	6.8	1080	24	ABI99537	Mouse ischaemic co
	44	62.6	6.8	654	23	AAS68925	DNA encoding novel
c	45	62.4	6.8	575	22	ABA63148	Human foetal liver

ALIGNMENTS

RESULT 1
AAA72432
ID AAA72432 standard; cDNA; 1048 BP.
XX
AC AAA72432;
XX
DT 19-DEC-2000 (first entry)
XX
DE Human nucleic acid-binding protein NuABP-51 cDNA.
XX
KW Human nucleic acid-binding protein; NuABP; agonist; antagonist; EST;
KW expressed sequence tag; drug screening; recombinant expression; antibody;
KW reproductive disorder; infertility; immunological disorder;
KW neurological disorder; cell proliferative disorder; cancer; tumour; ss.
XX
OS Homo sapiens.
XX
PN WO200044900-A2.
XX
PD 03-AUG-2000.
XX
PF 28-JAN-2000; 2000WO-US02237.
XX
PR 29-JAN-1999; 99US-0117904.
PR 29-JAN-1999; 99US-0117905.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Lal P, Hillman JL, Yue H, Azimzai Y, Lu AMD, Baughn MR;
PI Tran B, Shih LL, Au-Young JL;
XX
DR WPI; 2000-499332/44.

```
DR   P-PSDB; AAB21047.
XX
PT   Novel nucleic acid binding proteins, used to identify agonists and
PT   antagonists of them, for the treatment of reproductive, immunological,
PT   neurological and cell proliferative disorders including cancer -
XX
PS   Claim 4; Page 178; 180pp; English.
XX
CC   Sequences AAA72382-A72436 represent cDNAs encoding novel human nucleic
CC   acid-binding proteins (NuABPs; AAB20997-B21051). These cDNAs were
CC   produced by extension from an appropriate EST (expressed sequence
CC   tag) using primers designed using the EST. The invention also relates
CC   to expression constructs, host cells and transgenic organisms comprising
CC   a human NuABP nucleic acid, recombinant production of the human NuABPs,
CC   and antibodies against the human NuABPs, and also to methods of
CC   screening modulators of human NuABP activity or expression. The human
CC   NuABPs, and their agonists and antagonists are used to treat diseases
CC   associated with overexpression or underexpression of functional NuABPs.
CC   Human NuABP proteins and nucleotides, and NuABP agonists and antagonists
CC   can be used to diagnose, treat and prevent reproductive, immunological,
CC   neurological and cell proliferative disorders. Reproductive disorders
CC   that may be treated using compositions of the invention include
CC   infertility, endometriosis, disruptions of the menstrual cycle and
CC   disruptions of spermatogenesis. Immunological disorders that may be
CC   treated include AIDS, allergies, and autoimmune disorders such as
CC   multiple sclerosis, rheumatoid arthritis, diabetes and systemic lupus
CC   erythematosus. Neurological disorders that may be treated include
CC   epilepsy, neurodegenerative conditions such as Alzheimer's disease and
CC   Parkinson's disease, prion diseases such as Creutzfeldt-Jakob disease,
CC   and mental disorders such as schizophrenia. Cell proliferative disorders
CC   that may be treated include a wide variety of cancers, and also
CC   arteriosclerosis, atherosclerosis, cirrhosis and psoriasis.
XX
SQ   Sequence 1048 BP; 256 A; 289 C; 338 G; 165 T; 0 other;

  Query Match             92.9%;  Score 858;  DB 21;  Length 1048;
  Best Local Similarity   100.0%;  Pred. No. 3.6e-196;
  Matches  858;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CAGCCCGCTTCTCTGCCCGGAGCCATGAATCTCAGTAGCGCCAGTAGCACGGAGGAAAAG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     178 CAGCCCGCTTCTCTGCCCGGAGCCATGAATCTCAGTAGCGCCAGTAGCACGGAGGAAAAG 237

Qy      61 GCAGTGACGACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGACTTGGACCTTC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     238 GCAGTGACGACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGACTTGGACCTTC 297

Qy     121 AGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCATACGATTTGCTTGGGG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     298 AGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCATACGATTTGCTTGGGG 357

Qy     181 GAGAAAGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCAGCAAACCAGGAGGAC 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     358 GAGAAAGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCAGCAAACCAGGAGGAC 417

Qy     241 AAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCCCATGGTCTCC 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     418 AAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCCCATGGTCTCC 477

Qy     301 ATGGTAGGAGTGCAGCTTTCTCCCCCAGTTACTTTCCAGCTCCGGGCTGGCTCAGGACCC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     478 ATGGTAGGAGTGCAGCTTTCTCCCCCAGTTACTTTCCAGCTCCGGGCTGGCTCAGGACCC 537

Qy     361 GTGTTCCTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     538 GTGTTCCTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAG 597

Qy     421 GAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGAT 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     598 GAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGAT 657

Qy     481 ATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAG 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     658 ATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAG 717

Qy     541 GCGAGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTT 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     718 GCGAGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTT 777

Qy     601 AGAGACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     778 AGAGACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTC 837

Qy     661 AAGAAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTG 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     838 AAGAAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTG 897

Qy     721 CTGCAGGCACAGGGTGCCCCTGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTG 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     898 CTGCAGGCACAGGGTGCCCCTGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTG 957

Qy     781 TTGCGGGGGCAACATGAGAGCCCCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGT 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     958 TTGCGGGGGCAACATGAGAGCCCCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGT 1017

Qy     841 GAAGAGCCCCACCTCGGG 858
           ||||||||||||||||||
Db    1018 GAAGAGCCCCACCTCGGG 1035


RESULT 2
AAD00296
ID   AAD00296 standard; cDNA; 1019 BP.
XX
AC   AAD00296;
XX
DT   05-SEP-2000  (first entry)
XX
DE   Mouse oocyte-specific O1-236 cDNA clone.
XX
KW   Oocyte-specific; ovary; O1-236; mouse; Npm2; gynaecological; treatment;
KW   nucleoplasmin; cell proliferative disorder; cell degenerative disorder;
KW   contraceptive; ovulation; signalling pathway; human infertility; cancer;
KW   screen; modulator; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             156..779
FT                   /*tag=  a
FT                   /product= "Mouse oocyte-specific protein, O1-236"
FT                   /note= "Identical to mouse Npm2 protein with the
FT                   exception of one residue"
XX
PN   WO200024755-A1.
XX
PD   04-MAY-2000.
XX
PF   28-OCT-1999;   99WO-US25209.
XX
PR   28-OCT-1998;   98US-0106020.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI   Matzuk MM,  Wang P;
XX
DR   WPI; 2000-350684/30.
DR   P-PSDB; AAY70950.
XX
PT   O1-180, O1-184 and O1-236 polypeptides and nucleic acids encoding them,
PT   useful for evaluating potential contraceptives to block ovulation in a
PT   reversible manner  -
XX
PS   Claim 22; Fig 5; 54pp; English.
XX
CC   The present sequence is the mouse oocyte-specific gene Npm2, complete
```

CC cDNA clone O1-236. It is derived from mouse 2-cell embryo cDNA library
CC and expressed in the oocytes of intermediate size type 3a follicles and
CC all type 3b follicles. This clone is used to screen and identify the
CC mouse Npm2 gene. It is homologous to Xenopus laevis nucleoplasmin (Xnpm2)
CC expressed exclusively in eggs. It provides in vitro and in vivo reagents
CC for studying ovarian development and function. This sequence has
CC gynaecological and contraceptive activity. Agents which modulate O1-180,
CC O1-184 and O1-236 may be used to treat cell proliferative or degenerative
CC disorders, associated with abnormal expression of these ovary specific
CC genes. This ovary-specific sequence can be used as reagents to evaluate
CC potential contraceptives, to block ovulation in a reversible manner.
CC It is also used to screen for genetic mutations in signalling pathways,
CC that are associated with some forms of human infertility or
CC gynaecological cancers.
XX
SQ Sequence 1019 BP; 334 A; 243 C; 265 G; 177 T; 0 other;

Query Match 37.1%; Score 342.8; DB 21; Length 1019;
Best Local Similarity 66.4%; Pred. No. 1.5e-72;
Matches 603; Conservative 0; Mismatches 262; Indels 43; Gaps 6;

```
Qy      1 CAGCCCGCTTCTCTGCCCGGAGCCATGAATCTCAGTAGCGCCAGTAGCACGGAGGAAAAG 60
          ||  |   ||   | |||       | ||||| || |    ||| |||| |||  |    ||||
Db    132 CACCAAGCCGCCCTGTAAATCGACATGAGTCGCCACAGCACCAGCAGCGTGACCGAAACC 191

Qy     61 GCAGTGACGACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGACTTGGACCTTC 120
           |||   |   | | ||||||||||   | || |||| |||||| | || ||||||  |||||
Db    192 ACAGCAAAAAACATGCTCTGGGGTAGTGAACTCAATCAGGAAAAGCAGACTTGCACCTTT 251

Qy    121 AGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCATACGATTTGCTTGGGG 180
          |||  |||    |||  |||| | ||||| |   || |||||    ||||| ||| |||||
Db    252 AGAGGCCAAGGCGAGAAGAAGGACAGCTGTAAACTCTTGCTCAGCACGATCTGCCTGGGG 311

Qy    181 GAGAAAGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCAGCAAACCAGGAGGAC 240
          |||||||||||||||||| || | || |||||  ||||  |||             ||||
Db    312 GAGAAAGCCAAAGAGGAGGTGAACCGTGTGGAAGTCCTCTCCC-----------AGGAA 359

Qy    241 AAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCCCATGGTCTCC 300
             |  |  |  ||  |||| |||||  | ||  |||| |||||||| ||||||||| |
Db    360 GGCAGAAAACCACCAATCACTATTGCTACGCTGAAGGCATCAGTCCTGCCCATGGTCACT 419

Qy    301 ATGGTAGGAGTGCAGCTTTCTCCCCCAGTTACTTTCCAGCTCCGGGCTGGCTCAGGACCC 360
           ||  |||  |  |||||||||| ||||| ||||| | |||| || ||||||||||||
Db    420 GTGTCAGGTATAGAGCTTTCTCCTCCAGTAACTTTTCGGCTCAGGACTGGCTCAGGACCT 479

Qy    361 GTGTTCCTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAG 420
          |||||||||||||||| |||| |||||||  | || ||||| |||||||| || || |||
Db    480 GTGTTCCTCAGTGGCCTGGAATGTTATGAGACTTCGGACCTGACCTGGGAAGATGACGAG 539

Qy    421 GAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGAT 480
          |||           |||||||||||||| |||||||| ||||| ||||||||||||||
Db    540 GAA-----------GAGGAGGAAGAGGAGGAGGAAGAGGATGAAGATGAGGATGCAGAT 587

Qy    481 ATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAG 540
          |||||  || ||||||   |  |||||||||||||||||||| |||  ||||||||||||
Db    588 ATATCGCTAGAGGAG---ATACCTGTCAAACAAGTCAAAAGGGTGGCTCCCCAGAAGCAG 644

Qy    541 GCGAGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAA------TAAGAGCC 594
            |||| | || |||||||| ||| |||||||||||| || ||||      | ||  ||
Db    645 ATGAGCATAGCAAAGAAAAAGAAGGTGGAAAAAGAAGAGGATGAAACAGTAGTGAGGCCC 704

Qy    595 AGCGTTAGAGACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCA 654
          |||  |   |||||||| ||   ||| ||||  ||| | ||| ||||||| |||||||||
Db    705 AGCCCTCAGGACAAGAGTCCCTGGAAGAAGGAGAAATCTACACCCAGAGCAAAGAAGCCA 764

Qy    655 GGATTCAAGAAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGG 714
          |    ||||||||||     ||           |   || ||| |  ||   |||
Db    765 GTGACCAAGAAATGA--CCTCATCTTAGCATCTTCTGCGTCCAAGGCAGGATGTCCAGCA 822

Qy    715 GCTGTGCTGCAGGCACAGGGTGCCCCTGTCCAGCCCCTCCACCTGTGTCTGAATGCAACA 774
          ||||||  |          |||||   |||||||||| |||||   ||||||||| || |
Db    823 GCTGTGTT--------TTGGTGCAGGTGTCCAGCCCCACCACCCTAGTCTGAATGTAATA 874
```

```
Qy    775 GGGGTGTTGCGGGGGCAACATGAGAGCCCCTCACCCCCAACTCTCCACTTTCAGGAGGCC 834
            ||  ||    |  |  | | ||  | ||   |  ||       |     |||  || |
Db    875 AGGTGGTGTGGCTGTAACCCTGTAACCCAGCCCTCCAGTTTCCGGAGGTTTTTGGTGAAG 934

Qy    835 CCCAGTGAAGAGCCCCACCTCGGGGTCACAATAAAGTTGCCTGGTCAGGAAAAAAAAAAA 894
            |    |  |    | ||| |||     |  ||| |||| || ||||||||||||||||
Db    935 AGCCCCCAGCAAGTTCGCCTAGGGCCACAATAAAATTTGCATGATCAGGAAAAAAAAAAA 994

Qy    895 AAAAAAAA 902
            ||||||||
Db    995 AAAAAAAA 1002
```

RESULT 3
AAD23440
ID AAD23440 standard; cDNA; 423 BP.
XX
AC AAD23440;
XX
DT 26-FEB-2002 (first entry)
XX
DE Human lung tumour-specific 54589.1 cDNA.
XX
KW Human; lung tumour protein; immunostimulant; cytostatic; gene therapy;
KW antisense-therapy; vaccine; immune response; lung cancer; 54589.1; ss.
XX
OS Homo sapiens.
XX
PN WO200172295-A2.
XX
PD 04-OCT-2001.
XX
PF 28-MAR-2001; 2001WO-US09991.
XX
PR 29-MAR-2000; 2000US-0538037.
PR 05-JUN-2000; 2000US-0588937.
PR 18-AUG-2000; 2000US-0640878.
PR 22-SEP-2000; 2000US-234517P.
PR 01-NOV-2000; 2000US-0704512.
PR 14-DEC-2000; 2000US-0738973.
XX
PA (CORI-) CORIXA CORP.
XX
PI Reed SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indirias CY;
PI Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;
XX
DR WPI; 2001-639201/73.
XX
PT New human lung-specific polynucleotides and polypeptides for the
PT diagnosis and treatment of disease e.g. lung cancer -
XX
PS Claim 1; Page 302; 378pp; English.
XX
CC The invention relates to isolated lung tumour-specific proteins and
CC their corresponding cDNA molecules. Lung tumour-specific proteins and
CC their antigen-presenting cells are useful for stimulating and/or
CC expanding T cells specific for a tumour protein, and for inhibiting
CC the development of cancer. The invention also relates to a composition
CC useful for stimulating an immune response, and for treating cancer. The
CC lung tumour specific oligonucleotide is useful in gene therapy and for
CC diagnosis, detection and treatment of lung cancer. The present sequence
CC is a cDNA encoding human lung tumour-specific protein.
XX
SQ Sequence 423 BP; 84 A; 138 C; 125 G; 76 T; 0 other;

Query Match 36.5%; Score 337; DB 23; Length 423;
Best Local Similarity 100.0%; Pred. No. 2.7e-71;
Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 CAGCCCGCTTCTCTGCCCGGAGCCATGAATCTCAGTAGCGCCAGTAGCACGGAGGAAAAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     87 CAGCCCGCTTCTCTGCCCGGAGCCATGAATCTCAGTAGCGCCAGTAGCACGGAGGAAAAG 146
```

```
Qy      61 GCAGTGACGACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGACTTGGACCTTC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     147 GCAGTGACGACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGACTTGGACCTTC 206

Qy     121 AGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCATACGATTTGCTTGGGG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     207 AGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCATACGATTTGCTTGGGG 266

Qy     181 GAGAAAGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCAGCAAACCAGGAGGAC 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     267 GAGAAAGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCAGCAAACCAGGAGGAC 326

Qy     241 AAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCCCATGGTCTCC 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     327 AAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCCCATGGTCTCC 386

Qy     301 ATGGTAGGAGTGCAGCTTTCTCCCCCAGTTACTTTCC 337
           |||||||||||||||||||||||||||||||||||||
Db     387 ATGGTAGGAGTGCAGCTTTCTCCCCCAGTTACTTTCC 423
```

RESULT 4
ABK83560/c
ID ABK83560 standard; cDNA; 86574 BP.
XX
AC ABK83560;
XX
DT 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #131.
XX
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
XX
PN WO200228999-A2.
XX
PD 11-APR-2002.
XX
PF 03-OCT-2001; 2001WO-US30821.
XX
PR 03-OCT-2000; 2000US-237189P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
DR WPI; 2002-435328/46.
XX
PT Detecting granulocyte activation by detecting differential expression
PT of genes associated with granulocyte activation, which serves as
PT diagnostic markers that is useful for monitoring disease states and
PT drug toxicity -
XX
PS Claim 1; SEQ ID No 131; 114pp; English.
XX
CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) GA by contacting GC with an agent
CC that alters the expression of at least one gene in Gs; (2) screening (M3)
CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease using the
CC gene expression profile; (3) detecting (M4) an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease, by detecting the
CC level of expression in a sample of the tissue of gene(s) from Gs, where
CC the level of expression of the gene is indicative of inflammation;
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC an allergic response in a subject, exposure of a subject to a pathogen
CC or sterile inflammatory disease, by contacting a tissue having
CC inflammation with an agent that modulates the expression of gene(s)
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC modulating GA; M3 is useful for screening an agent capable of modulating
CC GCA preferably in an inflammation in a tissue; M4 is useful for
CC detecting an inflammation (especially chronic) in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, ARDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC periodontal disease; also bacterial infection, viral infection,
CC parasitic infection, protozoal infection, fungal infection and M5 is
CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 86574 BP; 22071 A; 20398 C; 21552 G; 22553 T; 0 other;

```
  Query Match           18.6%;  Score 171.6;  DB 24;  Length 86574;
  Best Local Similarity  71.9%;  Pred. No. 7.4e-31;
  Matches  299;  Conservative    0;  Mismatches   74;  Indels   43;  Gaps    4;

Qy     428 AAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATATATCTC 487
           | ||  | | ||||||  |||||  |||||||||||||||||  |||||||||    ||||
Db   16224 AGGGAAAAGGGGAGGAGGAGGAGGAAGATGATGAGGATAAGGATGCAGAAGGGTCTC 16165

Qy     488 TGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCG 547
           |||||  ||  |  |  ||||||     ||||||||||||||| ||||||||||||  |||||
Db   16164 TGGAGAAGGAGACCCCTGT----CAAGTCAAAAGGCTGGCGCCCCAGAAGCAGACGAGCT 16109

Qy     548 TGGCTAAGAAAAAAAAAGCTGGAAAAAGAAGA----------AGAGGAAATAAGAGC 593
           | ||  ||  ||||||  |    | |||  |  ||  |           |||||||  | |||  |
Db   16108 TTGCGAAAAAAAAAAGAAAAGAAAAGAAAAAAAAAACTGGAAAAAGAGAGGAAGTGAGACC 16049

Qy     594 CAGCGTTAGAGACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCC 653
            ||  |||||||  ||||||||||   ||||  ||||||||   ||||  |||  |
Db   16048 TAGTGTTAGAGGCAAGAGCCCTTGGAAATAGGCCAAACACACATTCAGGCC--------- 15998

Qy     654 AGGATTCAAGAAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTG 713
                 ||||||||||||||||  |||||||  ||  ||  ||||||  ||||   |||  |||
Db   15997 ------CAAGAAATGAGGAGCCATGCCTTGGAGGTCATGGTGCAGAGTGAACCTGCCCCT 15944

Qy     714 GGCTGTGCTGCAGGCACAGGGTGCCCCTGTCCAGCCCCTCCACCTGTGTCTGAATGCAAC 773
           ||||||||||  ||  ||||  |||           ||||||||  ||||||||||||   ||
Db   15943 GGCTGTGCTGTGGGTGCAGGATGC----------CCCCTCCATCTGTGTCTGAATATGAC 15894

Qy     774 AGGGGTGTTGCGGGGGCAACATGAGAGCCCCTCACCCCCAACTCTCCACTTTCAGG 829
           ||||||||||  ||||||||||  |  ||  ||||||||||  ||||||||   |  |||
Db   15893 AGGGGTGTTGTGGGGGCAACACTAAAGTCCCTCACCCCCGACTCTCCAGGTGGAGG 15838
```

RESULT 5
AAD00297
ID AAD00297 standard; DNA; 1481 BP.
XX
AC AAD00297;
XX
DT 05-SEP-2000 (first entry)
XX
DE Mouse ovary-specific Npm2 gene.
XX
KW Oocyte-specific; ovary; Npm2; nucleoplasmin; mouse; mammalian ortholog;

```
KW   chromosome 14; human chromosome 8p21; contraceptive; gynaecology; cancer;
KW   cell proliferative disorder; cell degenerative disorder; ovulation;
KW   modulator; human infertility; signalling pathway; screen; treatment; ds.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..150
FT                   /*tag= a
FT   exon            151..199
FT                   /*tag= b
FT                   /number= 1
FT   intron          200..230
FT                   /*tag= c
FT                   /number= 1
FT   misc_feature    215
FT                   /*tag= d
FT                   /note= "Corresponds to 314 missing nucleotides of
FT                   intron 1 between bases 214 and 216"
FT   exon            231..394
FT                   /*tag= e
FT                   /number= 2
FT   CDS             337..1177
FT                   /*tag= f
FT                   /product= "Mouse ovary-specific Npm2 protein"
FT                   /note= "Coding region is interrupted by 8 introns"
FT   intron          395..425
FT                   /*tag= g
FT                   /number= 2
FT   misc_feature    410
FT                   /*tag= h
FT                   /note= "Corresponds to 105 missing nucleotides of
FT                   intron 2 between bases 409 and 411"
FT   exon            426..511
FT                   /*tag= i
FT                   /number= 3
FT   misc_feature    473
FT                   /*tag= j
FT                   /note= "The 'T' is replaced with 'G' in the cDNA"
FT   intron          512..542
FT                   /*tag= k
FT                   /number= 3
FT   misc_feature    527
FT                   /*tag= l
FT                   /note= "Corresponds to 63 missing nucleotides of
FT                   intron 3 between bases 526 and 528"
FT   exon            543..656
FT                   /*tag= m
FT                   /number= 4
FT   intron          657..687
FT                   /*tag= n
FT                   /number= 4
FT   misc_feature    672
FT                   /*tag= o
FT                   /note= "Corresponds to 2771 (2.77kb) missing nucleotides
FT                   of intron 4 between bases 671 and 673"
FT   exon            688..781
FT                   /*tag= p
FT                   /number= 5
FT   intron          782..812
FT                   /*tag= q
FT                   /number= 5
FT   misc_feature    797
FT                   /*tag= r
FT                   /note= "Corresponds to 1321 (1.32kb) missing nucleotides
FT                   of intron 5 between bases 796 and 798"
FT   exon            813..964
FT                   /*tag= s
FT                   /number= 6
FT   intron          965..995
FT                   /*tag= t
FT                   /number= 6
FT   misc_feature    980
FT                   /*tag= u
FT                   /note= "Corresponds to 157 missing nucleotides of
FT                   intron 6 between bases 979 and 981"
FT   exon            996..1036
FT                   /*tag= v
FT                   /number= 7
FT   intron          1037..1067
FT                   /*tag= w
FT                   /number= 7
FT   misc_feature    1052
FT                   /*tag= x
FT                   /note= "Corresponds to 471 missing nucleotides of
FT                   intron 7 between bases 1051 and 1053"
FT   exon            1068..1101
FT                   /*tag= y
FT                   /number= 8
FT   intron          1102..1132
FT                   /*tag= z
FT                   /number= 8
FT   misc_feature    1117
FT                   /*tag= aa
FT                   /note= "Corresponds to 63 missing nucleotides of
FT                   intron 8 between bases 1116 and 1118"
FT   exon            1133..1381
FT                   /*tag= ab
FT                   /number= 9
FT   misc_feature    1229
FT                   /*tag= ac
FT                   /note= "The 'C' is replaced with 'T' in the cDNA"
FT   polyA_signal    1361..1366
FT                   /*tag= ad
FT   3'UTR           1382..1481
FT                   /*tag= ae
XX
PN   WO200024755-A1.
XX
PD   04-MAY-2000.
XX
PF   28-OCT-1999;   99WO-US25209.
XX
PR   28-OCT-1998;   98US-0106020.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI   Matzuk MM,  Wang P;
XX
DR   WPI; 2000-350684/30.
DR   P-PSDB; AAY70951.
XX
PT   O1-180, O1-184 and O1-236 polypeptides and nucleic acids encoding them,
PT   useful for evaluating potential contraceptives to block ovulation in a
PT   reversible manner  -
XX
PS   Example 5; Fig 13; 54pp; English.
XX
CC   The present sequence is the mouse ovary-specific Npm2 gene, mapped to
CC   the middle of mouse chromosome 14. It shows linkage to D14Mit32, this
CC   region being syntenic to human chromosome 8p21. The clone O1-236 cDNA is
CC   used to screen and identify Npm2 gene. It is the mammalian ortholog of
CC   Xenopus laevis nucleoplasmin (Xnpm2) expressed exclusively in the eggs.
CC   It provides in vitro and in vivo reagents for studying ovarian
CC   development and function. This sequence has gynaecological and
CC   contraceptive activity. Agents which modulate O1-180, O1-184 and O1-236
CC   may be used to treat cell proliferative or degenerative disorders,
CC   associated with abnormal expression of these ovary specific genes. This
CC   ovary-specific sequence can be used as reagents to evaluate potential
CC   contraceptives, to block ovulation in a reversible manner. It is also
CC   used to screen for genetic mutations in signalling pathways, that is
CC   associated with some form of human infertility or gynaecological cancer.
XX
SQ   Sequence 1481 BP; 407 A; 379 C; 391 G; 296 T; 8 other;

  Query Match              14.2%;  Score 131.6;  DB 21;  Length 1481;
```

```
  Best Local Similarity   63.1%;  Pred. No. 7.8e-22;
  Matches  274;  Conservative    0;  Mismatches  119;  Indels   41;  Gaps    3;

Qy     160 CTTCATACGATTTGCTTGGGGGAGAAAGCCAAAGAGGAGATGCATCGCGTGGAGATCCTG 219
           |||  |  ||| ||| ||||||||||||||||||||||| || | || |||||  ||||
Db     534 CTTTCTCAGATCTGCCTGGGGGAGAAAGCCAAAGAGGAGGTGAACCGTGTGGAAGTCCTC 593

Qy     220 CCCCCAGCAAACCAGGAGGACAAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCC 279
            |||  | |  |   |       |  |      |       |  || ||| ||| | ||
Db     594 TCCCAGGAAGGCAGAAAACCACCAATCACTATTGCTACGCTGAAGGCATCAGTCCTGCCC 653

Qy     280 TCAGTCCTCCCCAT------------------GGTCTCCATGGTAGGAGTGCAGCTTTC 320
             ||    |   |                  |||| |  ||  |||  |  |||||||
Db     654 ATGGTGAGTCTTCTCTCCNAGAAGGGGGACACAGGTCACTGTGTCAGGTATAGAGCTTTC 713

Qy     321 TCCCCCAGTTACTTTCCAGCTCCGGGCTGGCTCAGGACCCGTGTTCCTCAGTGGCCAGGA 380
           ||| ||||| ||||| | |||| || ||||||||||||| |||||||||||||||| |||
Db     714 TCCTCCAGTAACTTTTCGGCTCAGGACTGGCTCAGGACCTGTGTTCCTCAGTGGCCTGGA 773

Qy     381 ACGTTATGAAGCATCAGACC------------------TAACCTGGGAGGAGGAGGAGG 421
           | ||||||     |   | |                  || | |||   |     ||
Db     774 ATGTTATGGTAAGTTGTAGCCTANGGCTACCCATTCCAGAGACTTCGGACCTGACCTGGG 833

Qy     422 AAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATA 481
           |||| || | ||| |||||||||||||| |||||||| ||||| ||||||||||||||||
Db     834 AAGATGACGAGGAAGAGGAGGAAGAGGAGGAGGAAGAGGATGAAGATGAGGATGCAGATA 893

Qy     482 TATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGG 541
           |||| || ||||||   |  |||||||||||||||||||| |||  ||||||||||||
Db     894 TATCGCTAGAGGAG---ATACCTGTCAAACAAGTCAAAAGGGTGGCTCCCCAGAAGCAGA 950

Qy     542 CGAGCGTGGCTAAG 555
           ||||  |  ||  |||
Db     951 TGAGCATAGCAAAG 964


RESULT 6
AAL01804
ID   AAL01804 standard; cDNA; 416 BP.
XX
AC   AAL01804;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen cDNA SEQ ID NO: 1805.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
```

```
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465570/50.
DR   P-PSDB; AAM95834.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen
PT   is used in preventing, treating or ameliorating a medical condition  -
XX
PS   Claim 1; SEQ ID NO 1805; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a coding sequence of the
CC   invention.
XX
SQ   Sequence 416 BP; 96 A; 105 C; 130 G; 80 T; 5 other;

  Query Match             7.8%;  Score 72.4;  DB 22;  Length 416;
  Best Local Similarity   50.0%;  Pred. No. 8.2e-08;
  Matches  203;  Conservative    1;  Mismatches  196;  Indels    6;  Gaps    1;

Qy      51 GGAGGAAAAGGCAGTGACGACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGAC 110
           | | ||  |||  |  ||    | | |  | ||||| ||||||       | |  ||  |
Db       2 GCACGAGGAGGGTGGTACTGGGGCGGTAAGAGGCTGTGAGCTCTCCGGCCACACCCGCTC 61

Qy     111 TTGGACCTTCAGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCATACGAT 170
           |  ||||| |       ||   ||||    |   |  :    | |||   ||    | | |
Db      62 CTTCACCTTTAAGGTAGAGGAAGAGGATGATGCGGASACGTGNCTGGCACTAACCATGCT 121

Qy     171 TTGCTTGGGGGAGAAAGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCAGCAAA 230
           ||| |    |||  |||||||| |||    ||   || ||  |  || |||
Db     122 CTGCCTCACCGAGGGAGCCAAAGACGAGTGTAATGTGGTAGAAGTTGTGGCCCGGA---- 177

Qy     231 CCAGGAGGACAAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCC 290
             | ||| |  ||  |    | ||| |  | |||   ||| ||   ||   ||  ||
Db     178 --ACCATGACCATCAGGAGATCGCAGTCCCTGTGGCCAACCTCAAGCTGTCCTGCCAACC 235

Qy     291 CATGGTCTCCATGGTAGGAGTGCAGCTTTCTCCCCCAGTTACTTTCCAGCTCCGGGCTGG 350
           |||| ||    |||  |   | |||||    || || || || ||||  ||   | | ||
Db     236 CATGCTCAGTCTGGATGACTTCCAGCTCCAACCACCTGTAACCTTCCGCCTGAAGTCGGG 295

Qy     351 CTCAGGACCCGTGTTCCTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGA 410
           || || || |||    ||| ||| | | | |    ||   |      |
Db     296 TTCTGGCCCTGTGCGGNTCANTGGGCGGCACCAGATTGTTACGATGAGCAATGATGTTTC 355

Qy     411 GGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAA 456
           |||||||||   || |||| ||| |||||    || || || |||
Db     356 TGAGGAGGAGAGCGAGGAAGAGGAAGAGGACANTGATGAGGAAGAA 401


RESULT 7
ABL97097
ID   ABL97097 standard; cDNA; 416 BP.
XX
AC   ABL97097;
XX
DT   21-JUN-2002  (first entry)
XX
DE   Human testicular antigen encoding cDNA SEQ ID NO: 765.
XX
KW   Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW   reproductive system disorder; urinary system disorder; gene therapy;
KW   cardiovascular disorder; respiratory disorder; neurological disorder;
KW   gastrointestinal disease; infection; cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200155317-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01329.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
```

PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483232/52.
XX
PT Nucleic acids encoding 973 human testicular antigen polypeptides,
PT useful for preventing, diagnosing and/or treating testicular cancer -
XX
PS Claim 1; SEQ ID NO 765; 766pp; English.
XX
CC The present invention provides the protein and coding sequences of 973
CC human testicular antigens, and fragments of their genomic sequences. The
CC sequences can be used in the treatment of cardiovascular, urinary system,
CC reproductive system, immune, respiratory, neurological and
CC gastrointestinal disorders, infections, and particularly cancer,
CC especially testicular cancers. The present sequence is a cDNA of the
CC invention.
XX
SQ Sequence 416 BP; 96 A; 105 C; 130 G; 80 T; 5 other;

Query Match 7.8%; Score 72.4; DB 23; Length 416;
Best Local Similarity 50.0%; Pred. No. 8.2e-08;
Matches 203; Conservative 1; Mismatches 196; Indels 6; Gaps 1;

```
Qy     51 GGAGGAAAAGGCAGTGACGACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGAC 110
          | | ||  |||  |  ||    | | |  | ||||| ||||||       | |  ||  |
Db      2 GCACGAGGAGGGTGGTACTGGGGCGGTAAGAGGCTGTGAGCTCTCCGGCCACACCCGCTC 61

Qy    111 TTGGACCTTCAGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCATACGAT 170
            |  ||||| |      ||   ||||   |   |  :    | |||   ||    | | |
```

```
Db      62 CTTCACCTTTAAGGTAGAGGAAGAGGATGATGCGGASACGTGNCTGGCACTAACCATGCT 121

Qy     171 TTGCTTGGGGGAGAAAGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCAGCAAA 230
           |||  |    |||  ||||||||| |||    ||   || ||  |  || |||
Db     122 CTGCCTCACCGAGGGAGCCAAAGACGAGTGTAATGTGGTAGAAGTTGTGGCCCGGA---- 177

Qy     231 CCAGGAGGACAAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCC 290
             |  | |||  |  ||  |    | |||  |  | |||   |||  ||   ||   ||  ||
Db     178 --ACCATGACCATCAGGAGATCGCAGTCCCTGTGGCCAACCTCAAGCTGTCCTGCCAACC 235

Qy     291 CATGGTCTCCATGGTAGGAGTGCAGCTTTCTCCCCCAGTTACTTTCCAGCTCCGGGCTGG 350
           ||||  ||    |||  |   | |||||    || || || ||  ||||  ||   | | ||
Db     236 CATGCTCAGTCTGGATGACTTCCAGCTCCAACCACCTGTAACCTTCCGCCTGAAGTCGGG 295

Qy     351 CTCAGGACCCGTGTTCCTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGA 410
           || || || |||    ||| ||| | | | |    ||   |      |
Db     296 TTCTGGCCCTGTGCGGNTCANTGGGCGGCACCAGATTGTTACGATGAGCAATGATGTTTC 355

Qy     411 GGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAA 456
           |||||||||   || |||| ||| |||||    || || || |||
Db     356 TGAGGAGGAGAGCGAGGAAGAGGAAGAGGACANTGATGAGGAAGAA 401
```

RESULT 8
AAZ23891
ID AAZ23891 standard; DNA; 49999 BP.
XX
AC AAZ23891;
XX
DT 25-JAN-2000 (first entry)
XX
DE Murine LOBO genomic DNA fragment 1.
XX
KW LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW spondyloepiphysal dysplasia; achondroplasia; murine; ds.
XX
OS Mus musculus.
XX
PN WO9950284-A2.
XX
PD 07-OCT-1999.
XX
PF 26-MAR-1999; 99WO-EP02055.
XX
PR 27-MAR-1998; 98DE-1013799.
XX
PA (ROSE/) ROSENTHAL A.
XX
PI Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
XX
DR WPI; 1999-601320/51.
XX
PT Nucleic acids encoding proteins which influence bone development,
PT useful for treating and studying bone disorders -
XX
PS Example 3; Page 69-97; 391pp; German.
XX
CC This invention describes novel nucleic acids (I; designated LOBO (long
CC bones)) encoding proteins influencing bone development in mammals. The
CC proteins of the invention reduce and/or inactivate bone extension (i.e.
CC development), with exception of the skull and have osteopathic activity.
CC The nucleic acid molecules, proteins and antibodies can be used in
CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC and nucleic acid molecules, etc. are useful for production of transgenic
CC animals, especially a transgenic mouse for the study of diseases
CC associated with bone development, e.g. spondyloepiphysal dysplasia and
CC achondroplasia. This sequence encodes the murine LOBO protein described
CC in the method of the invention.
XX
SQ Sequence 49999 BP; 13210 A; 11814 C; 10825 G; 14150 T; 0 other;

Query Match 7.6%; Score 70.2; DB 20; Length 49999;
Best Local Similarity 53.5%; Pred. No. 1.3e-06;
Matches 147; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

```
Qy     377 AGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAGGAAGAAGAAGGGGAGG 436
           | |||    | ||||  |   ||   |     ||| || || ||||||||||||| ||| |
Db    7420 AAGAAGAGGAAGAAGAAGAGGAAGAAGAAGAGGAAGAAGAAGAGGAAGAAGAAGAGGAAG 7479

Qy     437 AGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATATATCTCTGGAGGAGC 496
           | || ||||| ||||| ||||| || || || || || || | |||   |      || ||
Db    7480 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 7539

Qy     497 AAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCGTGGCTAAGA 556
           ||       |   || |||   ||      | |    || || ||| |   | ||    ||
Db    7540 AAGAAGAAGAAGAAGAAGAAGAAGACGAGGAGGAGGAGGAGGAGGAGGGGGGGGGGA 7599

Qy     557 AAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTAGAGACAAGAGCCCTG 616
           | |  |||   ||| |||||| ||| | |   |      |  |  |  || ||
Db    7600 AGAGGAAGAAAGAAGAAGAAGGAGACGGAGAGAAGAAGAAGGAGAAGGAAAAAGAGAAGA 7659

Qy     617 TGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAG 651
           |||  ||||  || |  | ||  |  |  ||||||
Db    7660 AGAAGAAGGAGAAGGAGAAAGAGAAGGAGAAGAAG 7694
```

RESULT 9
AAZ23896
ID AAZ23896 standard; DNA; 49999 BP.
XX
AC AAZ23896;
XX
DT 25-JAN-2000 (first entry)
XX
DE Murine LOBO homologue genomic DNA fragment 2.
XX
KW LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW spondyloepiphysal dysplasia; achondroplasia; murine; ds.
XX
OS Mus musculus.
XX
PN WO9950284-A2.
XX
PD 07-OCT-1999.
XX
PF 26-MAR-1999; 99WO-EP02055.
XX
PR 27-MAR-1998; 98DE-1013799.
XX
PA (ROSE/) ROSENTHAL A.
XX
PI Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
XX
DR WPI; 1999-601320/51.
XX
PT Nucleic acids encoding proteins which influence bone development,
PT useful for treating and studying bone disorders -
XX
PS Example 3; Page 161-189; 391pp; German.
XX
CC This invention describes novel nucleic acids (I; designated LOBO (long
CC bones)) encoding proteins influencing bone development in mammals. The
CC proteins of the invention reduce and/or inactivate bone extension (i.e.
CC development), with exception of the skull and have osteopathic activity.
CC The nucleic acid molecules, proteins and antibodies can be used in
CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC and nucleic acid molecules, etc. are useful for production of transgenic
CC animals, especially a transgenic mouse for the study of diseases
CC associated with bone development, e.g. spondyloepiphysal dysplasia and
CC achondroplasia. This sequence encodes the murine LOBO protein described
CC in the method of the invention.
XX
SQ Sequence 49999 BP; 13135 A; 11787 C; 10868 G; 14209 T; 0 other;

Query Match 7.6%; Score 70.2; DB 20; Length 49999;
Best Local Similarity 53.5%; Pred. No. 1.3e-06;
Matches 147; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

```
Qy     377 AGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAGGAAGAAGAAGGGGAGG 436
           | |||    | |||| |   ||   |     ||| || || ||||||||||||| ||| |
Db    9405 AAGAAGAGGAAGAAGAAGAGGAAGAAGAAGAGGAAGAAGAAGAGGAAGAAGAAGAGGAAG 9464

Qy     437 AGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATATATCTCTGGAGGAGC 496
           | || ||||| ||||| ||||| || || || || || | |||   |      || ||
Db    9465 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 9524

Qy     497 AAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCGTGGCTAAGA 556
           ||      |   || |||   ||       | |    || || ||| |   | ||    ||
Db    9525 AAGAAGAAGAAGAAGAAGAAGAAGACGAGGAGGAGGAGGAGGAGGGGGGGGGGA 9584

Qy     557 AAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTAGAGACAAGAGCCCTG 616
           | |  |||   ||| |||||| ||| | |   |     |  |  |  || ||
Db    9585 AGAGGAAGAAAGAAGAAGAAGGAGACGGAGAGAAGAAGAAGGAGAAGGAAAAAGAGAAGA 9644

Qy     617 TGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAG 651
            ||| ||||  || |  | ||  |  |  ||||||
Db    9645 AGAAGAAGGAGAAGGAGAAAGAGAAGGAGAAGAAG 9679
```

RESULT 10
AAK67071
ID AAK67071 standard; DNA; 33923 BP.
XX
AC AAK67071;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:21883.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 21883; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 33923 BP; 9925 A; 8140 C; 7946 G; 7912 T; 0 other;

Query Match 7.5%; Score 69.2; DB 22; Length 33923;
Best Local Similarity 58.1%; Pred. No. 2e-06;
Matches 122; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```
Qy     377 AGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAGGAAGAAGAAGGGGAGG 436
           |||||     | || |      |||         |||  ||||||||||  | || | |||||
Db   22820 AGGAAGAGGAAGAGGAGGAGGACGAGGAGGAGGACAAGGAGGAGGACAAGGAGGAGGAGG 22879

Qy     437 AGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATATATCTCTGGAGGAGC 496
           ||||||| ||||| ||||| || || ||||| ||||| |  ||          |||||||
Db   22880 AGGAGGAGGAGGACGAGGAGGAGGAGGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGG 22939

Qy     497 AAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCGTGGCTAAGA 556
           | |     |  ||  | |  ||   |  |  |    ||||| ||| ||  | ||  |||
Db   22940 AGAAGGAGGAGAAGGAGGAGAAGGAGGAGAAGGAGGAGAAGGAGGACAAGGAGGAGAAGG 22999

Qy     557 AAAAAAAGCTGGAAAAAGAAGAAGAGGAAA 586
           |  | |||  ||| || || ||  |||| |
Db   23000 AGGAGAAGGAGGAGAAGGAGGAGAAGGAGA 23029
```

RESULT 11
AAS90676
ID AAS90676 standard; cDNA; 341 BP.
XX
AC AAS90676;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #26480.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG26489.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 26480; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.

```
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 341 BP; 174 A; 22 C; 124 G; 21 T; 0 other;

  Query Match             7.4%;  Score 68.8;  DB 23;  Length 341;
  Best Local Similarity   55.8%;  Pred. No. 5.6e-07;
  Matches  153;  Conservative    0;  Mismatches  117;  Indels    4;  Gaps    1;

Qy    388 GAAGCATCAGACCTAACCTGGGAGGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAG 447
          |||| |  ||    |     ||| ||||| ||||||||||||| ||| ||||||||||
Db     67 GAAGGAGGAGGAGAAGAAGAGGAAGAGGAAGAGGAAGAAGAAGAGGAAGAGGAGGAAGAA 126

Qy    448 GAAGAGGAAGATGATGAGGATGAGGATGCAGATATATCTCTGGAGGAGCAAAGCCCTGTC 507
          |||||||||||  || |||||  |||||  | |||  ||      || ||   |||
Db    127 GAAGAGGAAGAAGAAGAGGAAGAGGAAGAAGAAGTAGAAGAAGAAGAAAGAAG----AAG 182

Qy    508 AAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAGCTG 567
          ||       || | |  |  |    |||||  ||  ||    |   ||||| || |||  |
Db    183 AAGAGGAAGAAGAAGAAGAAGGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 242

Qy    568 GAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTAGAGACAAGAGCCCTGTGAAAAAGGCC 627
          | ||  |  |||| |||| ||||   || |  |||   |  ||       | ||
Db    243 AAGAAGAAGAAAGAAGAAAGAAGAAGAAGAGGAAGAAGAAGAAGAAGAAGAACAAGAAGA 302

Qy    628 AAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCA 661
          | |   | ||  |||    |||||  |  | | |
Db    303 AGAAGAAGAGGAAGAAGACAGAAGAAACAACTTA 336


RESULT 12
ABQ54697
ID   ABQ54697 standard; cDNA; 512 BP.
XX
AC   ABQ54697;
XX
DT   22-AUG-2002  (first entry)
XX
DE   Human ovarian antigen HTLEP21 cDNA, SEQ ID NO:577.
XX
KW   Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW   ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW   infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW   PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW   inflammatory condition; immune disorder; blood disorder;
KW   cardiovascular disorder; respiratory disorder; neurological disorder;
KW   gastrointestinal disorder; urinary system disorder; drug screening;
KW   gene therapy; chromosome mapping; forensic analysis;
KW   antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW   antiinflammatory; gynaecological; reproductive; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200200677-A1.
XX
PD   03-JAN-2002.
XX
PF   07-JUN-2001; 2001WO-US18569.
XX
PR   07-JUN-2000; 2000US-209467P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Birse CE,  Rosen CA;
XX
DR   WPI; 2002-147878/19.
DR   P-PSDB; ABP41620.
XX
PT   Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT   useful in the prevention, treatment and diagnosis of cancer (e.g.
PT   ovarian cancer), immune disorders, cardiovascular disorders and
PT   neurological diseases -
XX
PS   Claim 1; SEQ ID No 577; 2922pp; English.
XX
CC   The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC   ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC   encompasses polypeptides 90% identical and polynucleotides 95% identical
CC   to the sequences of the invention. The invention additionally relates to
CC   recombinant vectors and host cells comprising human ovarian antigen
CC   polynucleotides, antibodies against human ovarian antigens, and the use
CC   of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC   treating, prognosing or preventing various ovary and/or breast-related
CC   disorders. Such conditions include ovarian cancer and breast cancer, and
CC   metastatic tumours of ovarian or breast origin, reproductive system
CC   disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC   polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC   disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC   shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC   vaginitis), immune disorders (e.g., congenital and acquired
CC   immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC   blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC   respiratory disorders, neurological disorders, gastrointestinal disorders
CC   and urinary system disorders. Ovarian antigen polypeptides and
CC   polynucleotides may also be used in screening for compounds which
CC   modulate ovarian antigen expression or activity. The polynucleotides may
CC   further be used for gene therapy, chromosome mapping, in the
CC   identification of individuals and in forensic analysis, and the
CC   polypeptides may be used as food additives or to prepare antibodies
CC   useful in disease diagnosis, drug targeting and phenotyping. The present
CC   sequence represents cDNA encoding a human ovarian antigen of the
CC   invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 512 BP; 116 A; 137 C; 154 G; 99 T; 6 other;

  Query Match             7.3%;  Score 67.8;  DB 24;  Length 512;
  Best Local Similarity   50.4%;  Pred. No. 1.1e-06;
  Matches  207;  Conservative    5;  Mismatches  192;  Indels    7;  Gaps    2;

Qy     51 GGAGGAAAAGGCAGTGACGACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGAC 110
          | | ||  |||  |  ||    | | |  : ||||| ||||||       | |  ||  |
Db      2 GCACGAGGAGGGTGGTACTGGGGCGGTAASAGGCTGTGAGCTCTCCGGCCACACCCGCTC 61

Qy    111 TTGGACCTTCAGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCATACGAT 170
           |  ||||| |       ||   ||||    |   | |: :|  ||||   ||    | | |
Db     62 CTTCACCTTTAAGGTAGAGGAAGAGGATGATGCG-GMMRCGTGCTGGCACTAACCATGCT 120

Qy    171 TTGCTTGGGGGAGAAAGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCAGCAAA 230
           ||| |    |||  |||||||||| |||    ||   || ||  |  || |||
Db    121 CTGCCTCACCGAGGGAGCCAAAGACGAGTGTAATGTGGTAGAAGTTGTGGCCC------G 174

Qy    231 CCAGGAGGACAAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCC 290
            |  | ||| |  ||  |    | ||| |  | |||   ||| ||   ||   ||  ||
Db    175 GAACCATGACCATCAGGAGATCGCAGTCCCTGTGGCCAACCTCAAGCTGTCCTGCCAACC 234

Qy    291 CATGGTCTCCATGGTAGGAGTGCAGCTTTCTCCCCCAGTTACTTTCCAGCTCCGGGCTGG 350
          ||||  ||    |||  |   | |||||    || || || || ||||  ||   | | ||
Db    235 CATGCTCAGTCTGGATGACTTCCAGCTCCAACCACCTGTAACCTTCCGCCTGAAGTCGGG 294

Qy    351 CTCAGGACCCGTGTTCCTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGA 410
          :|| || |: |||    ||| ||| | | | |    ||   |      |
Db    295 YTCTGGCCMTGTGCGGATCACTGGGCGGCACCAGATTGTTACGATGAGCAATGATGTTTC 354

Qy    411 GGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGA 461
          |||||||||   || |||| ||| |||||    || || || |||| |||
Db    355 TGAGGAGGAGAGCGAGGAAGAGGAAGAGGACAGTGATGAGGAAGAAGTTGA 405


RESULT 13
AAS68926
ID   AAS68926 standard; cDNA; 291 BP.
XX
```

AC AAS68926;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #4730.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG04739.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 4730; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 291 BP; 129 A; 22 C; 112 G; 28 T; 0 other;

Query Match 7.2%; Score 66.4; DB 23; Length 291;
Best Local Similarity 54.0%; Pred. No. 2e-06;
Matches 136; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

```
Qy    362 TGTTCCTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAGG 421
          | | ||  | ||||||||   ||   |||        |  |     ||| |||||||| |
Db      8 TTTGCCATATTGGCCAGGCTAGTCTCGAACTCCTGACCTCAGGGGAGGAAGAGGAGGAAG 67

Qy    422 AAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATA 481
          | || |||| ||||||||| || || || |||||||| || || || ||||| | |||
Db     68 AGGAGGAAGAGGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAGGAAGAGG 127

Qy    482 TATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGG 541
          |     ||| ||| |       |   || |||   ||      |      |  |  | |
Db    128 AAGAGGAGGAAGAGGAGGAGGAAGAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 187
```

Qy 542 CGAGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTA 601
Db 188 AAGAAGAAGAAGAAGAAGAGGAAGAGGAAGAAGAAGAAGAAGAAGAAATAGCTTTAGGCG 247

Qy 602 GAGACAAGAGCC 613
Db 248 GAGATCAAATCC 259

RESULT 14
AAS68538
ID AAS68538 standard; cDNA; 1680 BP.
XX
AC AAS68538;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #4342.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG04351.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 4342; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1680 BP; 574 A; 286 C; 485 G; 335 T; 0 other;

Query Match 7.1%; Score 65.8; DB 23; Length 1680;

```
  Best Local Similarity   58.4%;  Pred. No. 5e-06;
  Matches  115;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy     408 GGAGGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGA 467
           |||||| |||||||| ||||||| ||||||||| || ||||| ||||||||  | |||||
Db    1476 GGAGGAAGAGGAGGAGGAAGAAGAGGAGGAGGAAGATGAGGAGGAGGAAGACAAGGAGGA 1535

Qy     468 TGAGGATGCAGATATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGT 527
           |||||| | ||| |       |||||||  ||         |||  | |   |   |  ||
Db    1536 TGAGGAGGAAGACAAGGAGAAGGAGGAGGAAGAAGACAACAAGGAGGAGGAGGAGGAGGA 1595

Qy     528 GCCCCAGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAAT 587
               | || || ||| |   |  |   | ||  |        | || | ||| ||  | |
Db    1596 AGACAAGGAGAAGGAGGAGGAAGAAGACAAGGAGGGGAGAAGGAAGACAAGGAGGAGGAGG 1655

Qy     588 AAGAGCCAGCGTTAGAG 604
           ||||    | |   |||
Db    1656 AAGACAAGGAGGAGGAG 1672


RESULT 15
AAK82672/c
ID   AAK82672 standard; DNA; 8165 BP.
XX
AC   AAK82672;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:37484.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
```

```
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 37484; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 8165 BP; 1595 A; 2353 C; 2025 G; 2192 T; 0 other;

  Query Match             7.1%;  Score 65.2;  DB 22;  Length 8165;
  Best Local Similarity   50.5%;  Pred. No. 1.2e-05;
  Matches  187;  Conservative    0;  Mismatches  178;  Indels    5;  Gaps    1;

Qy     307 GGAGTGCAGCTTTCTCCCCCAGTTACTTTCCAGCTCCGGGCTGGCTCAGGACCCGTGTTC 366
           | |||| | |    || | |   | || ||||||| | ||             |    |
Db    3706 GCAGTGAACCGAGATCACACCACTGCTCTCCAGCCCAGGTAACAGAGCCAGACTCCATCT 3647

Qy     367 CTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAGGAAGAA 426
           | ||      |  ||    |  ||| |  |||   ||    ||| || || ||||||||
Db    3646 CCCAAAAAAAAAAAAAAAAAAGTAAGAAGAAGAAAGAAGAAAGGAAGAAGAAGAGGAAGAG 3587

Qy     427 GAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATATATCT 486
           |||| |||||||||||| || ||||| || || || ||||| ||||| | ||| |
Db    3586 GAAGAGGAGGAGGAGGAGGAAGAAGAAGAGGAGGAGGAGGAAGAGGAAGAAGAGAAGAAG 3527

Qy     487 CTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGC 546
             | || || | ||        ||       || | |  |  |     |||  ||  ||
Db    3526 GAGAAGAAGAAGAGGAGGAGGAAGAGGAAGAAGAAGAGGAGGAGGAGGAATAAGAGGAAT 3467

Qy     547 GTGGCTAAGAAAAAAAAGCTG-----GAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTA 601
             |  ||||| || |||  |     ||| |||||||||| |||  |  ||     |
Db    3466 AAGAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 3407

Qy     602 GAGACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCA 661
           |||  |       | | |  | |  |||  | ||  |||   |||||   || |   |
Db    3406 AAGAAGAAGAAGAAGAGGAGGAAGAAGAAGAGAAAGAAAGAAGAAAGAAAAAAGAAAGAA 3347

Qy     662 AGAAATGAGG 671
           | | || |
Db    3346 GAAGAAGAAG 3337


Search completed: January 24, 2003, 05:27:29
Job time : 443 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 24, 2003, 03:53:00 ; Search time 2237 Seconds
(without alignments)
6689.598 Million cell updates/sec

Title: US-09-844-864-16
Perfect score: 924
Sequence: 1 cagcccgcttctctgcccgg.....tttgcggccgcaagcttatg 924

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	517	56.0	517	9	AI391464	AI391464 tf96g05.x
c	2	455	49.2	455	9	AI016313	AI016313 ot78a10.s
	3	433	46.9	433	13	BM055230	BM055230 ie92b10.y
c	4	432.2	46.8	451	13	BM054976	BM054976 ie92b10.x
	5	418.6	45.3	1324	14	BM809804	BM809804 AGENCOURT
	6	414.4	44.8	854	14	BQ878692	BQ878692 AGENCOURT
	7	410.6	44.4	915	14	BQ684634	BQ684634 AGENCOURT
	8	405.8	43.9	872	14	BQ878312	BQ878312 AGENCOURT
c	9	405	43.8	405	12	BF594409	BF594409 7i05a05.x
c	10	397.4	43.0	399	10	AW731946	AW731946 ba03e06.x
	11	389.6	42.2	935	14	BQ684424	BQ684424 AGENCOURT
	12	378.4	41.0	966	14	BQ682257	BQ682257 AGENCOURT
c	13	357.8	38.7	361	10	AW002370	AW002370 wu61a07.x
c	14	357.8	38.7	361	12	BF592761	BF592761 7j94b04.x
c	15	357.8	38.7	361	12	BF057162	BF057162 7k16g07.x
	16	352.4	38.1	377	12	BF804737	BF804737 CM0-CI009
c	17	350.8	38.0	378	12	BF223902	BF223902 7q36b08.x
	18	350.6	37.9	1047	14	BQ888606	BQ888606 AGENCOURT
	19	280.4	30.3	491	13	BI913169	BI913169 603179964
	20	234.2	25.3	636	10	BB559308	BB559308 BB559308
	21	178.4	19.3	580	13	BI439005	BI439005 ic25d05.y
	22	158.6	17.2	280	10	BB284820	BB284820 BB284820
	23	126.6	13.7	207	14	D81565	D81565 HUM173C12B
	24	103.2	11.2	548	12	BG515101	BG515101 dad59h10.
	25	103.2	11.2	623	9	AL682691	AL682691 AL682691
	26	103.2	11.2	658	9	AL782493	AL782493 AL782493
	27	102	11.0	566	9	AL645040	AL645040 AL645040
c	28	96.8	10.5	593	10	BE026525	BE026525 db27f07.x
	29	96.4	10.4	574	13	BJ094747	BJ094747 BJ094747
	30	96.4	10.4	600	14	BQ387680	BQ387680 NISC_mn25
c	31	96.4	10.4	645	13	BJ098416	BJ098416 BJ098416
	32	96	10.4	590	13	BI941416	BI941416 df88a05.y
c	33	95.8	10.4	555	12	BG021389	BG021389 df87a05.x
	34	94.4	10.2	380	10	BE506675	BE506675 db87d01.y
	35	94.2	10.2	522	12	BG886286	BG886286 dad51d02.
c	36	94.2	10.2	551	12	BG016384	BG016384 df88a05.x
	37	93.8	10.2	594	12	BF427228	BF427228 df87a05.y
c	38	93.4	10.1	564	10	BE678406	BE678406 df55d03.x
c	39	91.4	9.9	531	10	BE192089	BE192089 db87d01.x
c	40	90.2	9.8	460	12	BG515848	BG515848 dad59h10.
c	41	89	9.6	593	10	BE026442	BE026442 db26f09.x
	42	87.2	9.4	554	10	BE680412	BE680412 df79b03.y
	43	84.8	9.2	541	9	AL595075	AL595075 AL595075
c	44	83.6	9.0	549	14	BQ387679	BQ387679 NISC_mn25
	45	81	8.8	537	9	AL682539	AL682539 AL682539

ALIGNMENTS

RESULT 1
AI391464/c
LOCUS AI391464 517 bp mRNA linear EST 02-FEB-1999
DEFINITION tf96g05.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2107160 3' similar to contains element MER22 repetitive element ;, mRNA sequence.
ACCESSION AI391464
VERSION AI391464.1 GI:4217468
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 517)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever, M.D., Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco

```
            High quality sequence stop: 457.
FEATURES             Location/Qualifiers
     source          1. .517
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2107160"
                     /clone_lib="NCI_CGAP_CLL1"
                     /tissue_type="B-cell, chronic lymphotic leukemia"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTGCTTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT       70 a    168 c    121 g    158 t
ORIGIN

  Query Match             56.0%;  Score 517;  DB 9;  Length 517;
  Best Local Similarity   100.0%;  Pred. No. 6.1e-77;
  Matches  517;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     367 CTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAGGAAGAA 426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     517 CTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAGGAAGAA 458

Qy     427 GAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATATATCT 486
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     457 GAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATATATCT 398

Qy     487 CTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGC 546
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     397 CTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGC 338

Qy     547 GTGGCTAAGAAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTAGAGAC 606
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     337 GTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTAGAGAC 278

Qy     607 AAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAA 666
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     277 AAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAA 218

Qy     667 TGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTGCAG 726
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     217 TGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTGCAG 158

Qy     727 GCACAGGGTGCCCCTGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTGTTGCGG 786
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     157 GCACAGGGTGCCCCTGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTGTTGCGG 98

Qy     787 GGGCAACATGAGAGCCCCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGTGAAGAG 846
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      97 GGGCAACATGAGAGCCCCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGTGAAGAG 38

Qy     847 CCCCACCTCGGGGTCACAATAAAGTTGCCTGGTCAGG 883
           |||||||||||||||||||||||||||||||||||||
Db      37 CCCCACCTCGGGGTCACAATAAAGTTGCCTGGTCAGG 1


RESULT 2
AI016313/c
LOCUS       AI016313                455 bp    mRNA    linear   EST 27-AUG-1998
DEFINITION  ot78a10.s1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
            IMAGE:1622874 3' similar to contains element MER22 repetitive
            element ;, mRNA sequence.
ACCESSION   AI016313
VERSION     AI016313.1  GI:3230649
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1368   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 447.
FEATURES             Location/Qualifiers
     source          1. .455
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1622874"
                     /clone_lib="Soares_total_fetus_Nb2HF8_9w"
                     /dev_stage="8-9 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from pooled 8-9 week
                     (total) fetus material with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCTTAATTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       62 a    143 c    110 g    140 t
ORIGIN

  Query Match             49.2%;  Score 455;  DB 9;  Length 455;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-66;
  Matches  455;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     433 GAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATATATCTCTGGAG 492
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     455 GAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATATATCTCTGGAG 396

Qy     493 GAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCGTGGCT 552
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     395 GAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCGTGGCT 336

Qy     553 AAGAAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTAGAGACAAGAGC 612
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     335 AAGAAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTAGAGACAAGAGC 276

Qy     613 CCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAATGAGGA 672
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     275 CCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAATGAGGA 216

Qy     673 GCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTGCAGGCACAG 732
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     215 GCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTGCAGGCACAG 156

Qy     733 GGTGCCCCTGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTGTTGCGGGGGCAA 792
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     155 GGTGCCCCTGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTGTTGCGGGGGCAA 96

Qy     793 CATGAGAGCCCCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGTGAAGAGCCCCAC 852
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      95 CATGAGAGCCCCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGTGAAGAGCCCCAC 36

Qy     853 CTCGGGGTCACAATAAAGTTGCCTGGTCAGGAAAA 887
           ||||||||||||||||||||||||||||||||||||
Db      35 CTCGGGGTCACAATAAAGTTGCCTGGTCAGGAAAA 1


RESULT 3
```

```
BM055230
LOCUS       BM055230                433 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  ie92b10.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
            cDNA clone IMAGE:5674315 5', mRNA sequence.
ACCESSION   BM055230
VERSION     BM055230.1  GI:16812965
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 433)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Other_ESTs: ie92b10.x1
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu) This sequence now available from the IMAGE
            consortium, for clone orders contact: info@image.llnl.gov.
FEATURES              Location/Qualifiers
     source           1. .433
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:5674315"
                      /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                      /sex="Both"
                      /tissue_type="Islets of Langerhans"
                      /dev_stage="Adult"
                      /lab_host="DH10B"
                      /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
                      Site_2: Sal 1; Starting library constructed using
                      SuperScript Plasmid Library kit  (Life Technologies). cDNA
                      made by oligo-dT priming. Size-selected by  column
                      fractionation; average insert size 1.08 kb. Library was
                      amplified  once on solid support and plasmid DNA from
                      library was prepared. The  library DNA was normalized by
                      method #4 from Bonaldo, Lennon, and Soares  1996 Genome
                      Research 6:791-806; 0.5 microgram single-stranded library
                      plasmid DNA was mixed with 5 micrograms PCR product
                      representing library  inserts and hybridized to an Ecot of
                      20. Single-stranded (unhybridized)  plasmids were isolated
                      by hydroxyapatite chromatography and used to make  this
                      library."
BASE COUNT      137 a     98 c    143 g     55 t
ORIGIN

  Query Match           46.9%;  Score 433;  DB 13;  Length 433;
  Best Local Similarity  100.0%;  Pred. No. 6.6e-63;
  Matches  433;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     391 GCATCAGACCTAACCTGGGAGGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAA 450
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCATCAGACCTAACCTGGGAGGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAA 60

Qy     451 GAGGAAGATGATGAGGATGAGGATGCAGATATATCTCTGGAGGAGCAAAGCCCTGTCAAA 510
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GAGGAAGATGATGAGGATGAGGATGCAGATATATCTCTGGAGGAGCAAAGCCCTGTCAAA 120

Qy     511 CAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAGCTGGAA 570
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAGCTGGAA 180

Qy     571 AAAGAAGAAGAGGAAATAAGAGCCAGCGTTAGAGACAAGAGCCCTGTGAAAAAGGCCAAA 630
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAAGAAGAAGAGGAAATAAGAGCCAGCGTTAGAGACAAGAGCCCTGTGAAAAAGGCCAAA 240

Qy     631 GCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAATGAGGAGCCACGCCTTGGGGGGCA 690
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAATGAGGAGCCACGCCTTGGGGGGCA 300

Qy     691 CGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTGCAGGCACAGGGTGCCCCTGTCCAGCCC 750
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTGCAGGCACAGGGTGCCCCTGTCCAGCCC 360

Qy     751 CTCCACCTGTGTCTGAATGCAACAGGGGTGTTGCGGGGGCAACATGAGAGCCCCTCACCC 810
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CTCCACCTGTGTCTGAATGCAACAGGGGTGTTGCGGGGGCAACATGAGAGCCCCTCACCC 420

Qy     811 CCAACTCTCCACT 823
           |||||||||||||
Db     421 CCAACTCTCCACT 433


RESULT 4
BM054976/c
LOCUS       BM054976                451 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  ie92b10.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
            cDNA clone IMAGE:5674315 3', mRNA sequence.
ACCESSION   BM054976
VERSION     BM054976.1  GI:16812511
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 451)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu) This sequence now available from the IMAGE
            consortium, for clone orders contact: info@image.llnl.gov.
FEATURES              Location/Qualifiers
     source           1. .451
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:5674315"
                      /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
                      /sex="Both"
                      /tissue_type="Islets of Langerhans"
                      /dev_stage="Adult"
                      /lab_host="DH10B"
                      /note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
                      Site_2: Sal 1; Starting library constructed using
                      SuperScript Plasmid Library kit  (Life Technologies). cDNA
                      made by oligo-dT priming. Size-selected by  column
```

fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an Ecot of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 61 a 136 c 110 g 144 t
ORIGIN

Query Match 46.8%; Score 432.2; DB 13; Length 451;
Best Local Similarity 98.9%; Pred. No. 8.9e-63;
Matches 446; Conservative 0; Mismatches 3; Indels 2; Gaps 1;

```
Qy    445 GAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATATATCTCTGGAGGAGCAAAGCCCT 504
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    451 GAGGAAGAGGAAGATGATGAGGATGAGGATGCAGATATATCTCTGGAGGAGCAAAGCCCT 392

Qy    505 GTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAG 564
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    391 GTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAG 332

Qy    565 CTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTAGAGACAAGAGCCCTGTGAAAAAG 624
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    331 CTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTAGAGACAAGAGCCCTGTGAAAAAG 272

Qy    625 GCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAATGAGGAGCCACGCCTT-- 682
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    271 GCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAATGAGGAGCCACGCCATAA 212

Qy    683 GGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTGCAGGCACAGGGTGCCCCTG 742
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    211 GAGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTGCAGGCACAGGGTGCCCCTG 152

Qy    743 TCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTGTTGCGGGGGCAACATGAGAGCC 802
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    151 TCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTGTTGCGGGGGCAACATGAGAGCC 92

Qy    803 CCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGTGAAGAGCCCCACCTCGGGGTCA 862
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     91 CCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGTGAAGAGCCCCACCTCGGGGTCA 32

Qy    863 CAATAAAGTTGCCTGGTCAGGAAAAAAAAAA 893
          |||||||||||||||||||| |||||||||
Db     31 CAATAAAGTTGCCTGGTCAGGGAAAAAAAAA 1
```

RESULT 5
BM809804
LOCUS BM809804 1324 bp mRNA linear EST 05-MAR-2002
DEFINITION AGENCOURT_6581147 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5454702 5', mRNA sequence.
ACCESSION BM809804
VERSION BM809804.1 GI:19126627
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1324)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1950 row: e column: 07
High quality sequence stop: 357.
FEATURES Location/Qualifiers
source 1. .1324
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5454702"
/clone_lib="NIH_MGC_98"
/tissue_type="astrocytoma grade IV, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 383 a 90 c 706 g 60 t 85 others
ORIGIN

Query Match 45.3%; Score 418.6; DB 14; Length 1324;
Best Local Similarity 78.4%; Pred. No. 1.1e-60;
Matches 494; Conservative 0; Mismatches 134; Indels 2; Gaps 2;

```
Qy    106 CGGACTTGGACCTTCAGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCAT 165
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGGACTTGGACCTTCAGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCAT 60

Qy    166 ACGATTTGCTTGGGGGAGAAAGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCA 225
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACGATTTGCTTGGGGGAGAAAGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCA 120

Qy    226 GCAAACCAGGAGGACAAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTC 285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCAAACCAGGAGGACAAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTC 180

Qy    286 CTCCCCATGGTCTCCATGGTAGGAGTGCAGCTTTCTCCCCCAGTTACTTTCCAGCTCCGG 345
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTCCCCATGGTCTCCATGGTAGGAGTGCAGCTTTCTCCCCCAGTTACTTTCCAGCTCCGG 240

Qy    346 GCTGGCTCAGGACCCGTGTTCCTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACC 405
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCTGGCTCAGGACCCGTGTTCCTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACC 300

Qy    406 TGGGAGGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAG 465
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||| | ||
Db    301 TGGGAGGAGGAGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAANANGAANANGANNAN 360

Qy    466 GATGAGGATGCAGATATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTG 525
          || || || | | | | |     |||||| |||| | ||| ||| |||||| |
Db    361 GANGANGANGNANANANANNNNNGGAGGAAGAAAGNNNNGNAAAAAAAG-GAAAAGGGGG 419

Qy    526 GTGCCCCAGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAA 585
          | |    ||||| | | ||| | ||  || |||||||||  ||||||||||||||||||
Db    420 GGGNGNGAGAAGNANGNGAGGGNGGNNAAAAAAAAAAAAGNGGGAAAAAGAAGAAGAGGAA 479

Qy    586 ATAAG-AGCCAGCGTTAGAGACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGC 644
            ||| |   || |  ||||| |||||     | | || ||    ||          ||
Db    480 GNAAGAAAGGAGGGGGAGAGAAAAGAGGNGGGGGGAAGAGAGGNAAAGNNAGAGGGAAGA 539

Qy    645 CAAGAAGCCAGGATTCAAGAAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGG 704
              |||  ||||   ||     ||  ||          |||||   || |       ||
Db    540 GGGAAAGGAAGGAGGGAANGGGAGAAAAGAGGGAGGNGAGGGGGGGGGGGGGGNNAGGGA 599

Qy    705 CCTTCCCTGGGCTGTGCTGCAGGCACAGGG 734
                |||    |  | |||    |||
Db    600 GGAAAAGGGGGGGNNGGGGNAGGGGGGGGG 629
```

RESULT 6

```
BQ878692
LOCUS       BQ878692                854 bp    mRNA    linear   EST 16-AUG-2002
DEFINITION  AGENCOURT_8072876 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6087706
            5', mRNA sequence.
ACCESSION   BQ878692
VERSION     BQ878692.1  GI:22270700
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 854)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2324  row: l  column: 11
            High quality sequence stop: 560.
FEATURES             Location/Qualifiers
     source          1. .854
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6087706"
                     /clone_lib="NIH_MGC_112"
                     /tissue_type="melanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
BASE COUNT      181 a    247 c    253 g    151 t     22 others
ORIGIN

  Query Match             44.8%;  Score 414.4;  DB 14;  Length 854;
  Best Local Similarity   76.5%;  Pred. No. 6.7e-60;
  Matches  595;  Conservative    0;  Mismatches   16;  Indels  167;  Gaps    1;

Qy       1 CAGCCCGCTTCTCTGCCCGGAGCCATGAATCTCAGTAGCGCCAGTAGCACGGAGGAAAAG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     207 CAGCCCGCTTCTCTGCCCGGAGCCATGAATCTCAGTAGCGCCAGTAGCACGGAGGAAAAG 266

Qy      61 GCAGTGACGACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGACTTGGACCTTC 120
           ||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     267 GCAGTGACNNCCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGACTTGGACCTTC 326

Qy     121 AGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCATACGATTTGCTTGGGG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     327 AGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCATACGATTTGCTTGGGG 386

Qy     181 GAGAAAGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCAGCAAACCAGGAGGAC 240
           ||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||
Db     387 GAGAAAGCCAAAGAGGAGATGCATCGNNTGGAGATCCTGCCCCCAGCAAACCAGGAGGAC 446

Qy     241 AAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCCCATGGTCTCC 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     447 AAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCCCATGGTCTCC 506

Qy     301 ATGGTAGGAGTGCAGCTTTCTCCCCCAGTTACTTTCCAGCTCCGGGCTGGCTCAGGACCC 360
           ||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db     507 ATGGTAGGAGTGCAGCTTTCTCCCCCAGTTACTTTCCAGCTCNNNGCTGGCTCAGGACCC 566

Qy     361 GTGTTCCTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAG 420
           |||||||||||||||||||||||||||
Db     567 GTGTTCCTCAGTGGCCAGGAACGTTAT---------------------------------- 593

Qy     421 GAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGAT 480

Db     594 ---------------------------------------------------------- 593

Qy     481 ATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAG 540

Db     594 ---------------------------------------------------------- 593

Qy     541 GCGAGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTT 600
                           ||||||||||||||||||||||||||||||||||||||||||||||
Db     594 -------------GAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTT 639

Qy     601 AGAGACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTC 660
           |||||||||||||||||||||||||  |     | ||||||||||||||||||||||||||
Db     640 AGAGACAAGAGCCCTGTGAAAAAGGTCTTNNNCTCAGCCAGAGCCAAGAAGCCAGGATTC 699

Qy     661 AAGAAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTG 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     700 AAGAAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTG 759

Qy     721 CTGCAGGCACAGGGTGCCCCTGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGG 778
           ||||| |||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db     760 CTGCAAGCACAGGGTGCCCCTGTCCAGCCCCTCCACCTGTGTCTGAATGCAACGGGGG 817


RESULT 7
BQ684634
LOCUS       BQ684634                915 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8032849 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6091012
            5', mRNA sequence.
ACCESSION   BQ684634
VERSION     BQ684634.1  GI:21797313
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 915)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2333  row: f  column: 05
            High quality sequence stop: 652.
FEATURES             Location/Qualifiers
     source          1. .915
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6091012"
                     /clone_lib="NIH_MGC_112"
                     /tissue_type="melanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
BASE COUNT      197 a    283 c    277 g    153 t      5 others
```

```
ORIGIN

  Query Match           44.4%;  Score 410.6;  DB 14;  Length 915;
  Best Local Similarity 77.0%;  Pred. No. 2.8e-59;
  Matches  631;  Conservative    0;  Mismatches   17;  Indels  172;  Gaps    4;

Qy       1 CAGCCCGCTTCTCTGCCCGGAGCCATGAATCTCAGTAGCGCCAGTAGCACGGAGGAAAAG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     207 CAGCCCGCTTCTCTGCCCGGAGCCATGAATCTCAGTAGCGCCAGTAGCACGGAGGAAAAG 266

Qy      61 GCAGTGACGACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGACTTGGACCTTC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     267 GCAGTGACGACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGACTTGGACCTTC 326

Qy     121 AGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCATACGATTTGCTTGGGG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     327 AGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCATACGATTTGCTTGGGG 386

Qy     181 GAGAAAGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCAGCAAACCAGGAGGAC 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     387 GAGAAAGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCAGCAAACCAGGAGGAC 446

Qy     241 AAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCCCATGGTCTCC 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     447 AAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCCCATGGTCTCC 506

Qy     301 ATGGTAGGAGTGCAGCTTTCTCCCCCAGTTACTTTCCAGCTCCGGGCTGGCTCAGGACCC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     507 ATGGTAGGAGTGCAGCTTTCTCCCCCAGTTACTTTCCAGCTCCGGGCTGGCTCAGGACCC 566

Qy     361 GTGTTCCTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAG 420
           ||||||||||||||||||||||||||
Db     567 GTGTTCCTCAGTGGCCAGGAACGTTAT------------------------------------ 593

Qy     421 GAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGAT 480

Db     594 ------------------------------------------------------------ 593

Qy     481 ATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAG 540

Db     594 ------------------------------------------------------------ 593

Qy     541 GCGAGCGTGGCTAAGAAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTT 600
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     594 -------------GAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTT 639

Qy     601 AGAGACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTC 660
           |||||||||||||||| |||||||||||||||||||||||||||||  ||||||||
Db     640 AGAGACAAGAGCCCTGTG-AAAAGGCCAAAGCCACAGCCAGAGCCAAGAAAGCAGGATTC 698

Qy     661 AAGAAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTG 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     699 AAGAAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTG 758

Qy     721 CTGCAGGCACAGGGTGCCCCTGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGT- 779
           |||||||||||||||||||||||||||||  ||||||||||||    || ||||
Db     759 CTGCAGGCACAGGGTGCCCCTGTCCAGCCCCCTCACCTGTGTCTGAAATGCACNGGGGNT 818

Qy     780 --GTTGCGGGGGCAACATG-AGAGCCCCTCACCCCCAACT 816
             |   ||||||| |||| ||||||||||||||| | |
Db     819 GGTTGCNGGGGGCACCATGAAGAGCCCCTCACCCCCCAAT 858


RESULT 8
BQ878312
LOCUS       BQ878312                 872 bp    mRNA    linear   EST 16-AUG-2002
DEFINITION  AGENCOURT_8072779 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6087652
            5', mRNA sequence.
ACCESSION   BQ878312
VERSION     BQ878312.1  GI:22270320
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 872)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2324  row: j  column: 05
            High quality sequence stop: 548.
FEATURES             Location/Qualifiers
     source          1. .872
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6087652"
                     /clone_lib="NIH_MGC_112"
                     /tissue_type="melanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
BASE COUNT      192 a    265 c    259 g    145 t     11 others
ORIGIN

  Query Match           43.9%;  Score 405.8;  DB 14;  Length 872;
  Best Local Similarity 76.1%;  Pred. No. 1.8e-58;
  Matches  602;  Conservative    0;  Mismatches   20;  Indels  169;  Gaps    2;

Qy       1 CAGCCCGCTTCTCTGCCCGGAGCCATGAATCTCAGTAGCGCCAGTAGCACGGAGGAAAAG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     207 CAGCCCGCTTCTCTGCCCGGAGCCATGAATCTCAGTAGCGCCAGTAGCACGGAGGAAAAG 266

Qy      61 GCAGTGACGACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGACTTGGACCTTC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     267 GCAGTGACGACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGACTTGGACCTTC 326

Qy     121 AGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCATACGATTTGCTTGGGG 180
           |||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db     327 AGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCNNCATACGATTTGCTTGGGG 386

Qy     181 GAGAAAGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCAGCAAACCAGGAGGAC 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     387 GAGAAAGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCAGCAAACCAGGAGGAC 446

Qy     241 AAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCCCATGGTCTCC 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     447 AAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCCCATGGTCTCC 506

Qy     301 ATGGTAGGAGTGCAGCTTTCTCCCCCAGTTACTTTCCAGCTCCGGGCTGGCTCAGGACCC 360
           ||||||||||||||||||||||||    ||||||||||||||||||||||||||||||||
Db     507 ATGGTAGGAGTGCAGCTTTCTCCCNNNTTTACTTTCCAGCTCCGGGCTGGCTCAGGACCC 566

Qy     361 GTGTTCCTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGGAGGAG 420
           ||||||||||||||||||||||||||
Db     567 GTGTTCCTCAGTGGCCAGGAACGTTAT------------------------------------ 593

Qy     421 GAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATGCAGAT 480

Db     594 ------------------------------------------------------------ 593

Qy     481 ATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAG 540
```

```
Db     594 ------------------------------------------------------------ 593

Qy     541 GCGAGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTT 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     594 -------------GAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTT 639

Qy     601 AGAGACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     640 AGAGACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTC 699

Qy     661 AAGAAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTG 720
           |||||||||||||||||||||||||||||||||||||     | ||||||||||||||
Db     700 AAGAAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGGNNNCATTCCCTGGGCTGTG 759

Qy     721 CTGCAGGCACAGGGTGCCCCTGTCCAGCCCC--TCCACCTGTGTCTGAATGCAACAGGGG 778
           |||||||||||||||||||||||||||||||   ||||||||||||||||||||   |||
Db     760 CTGCAGGCACAGGGTGCCCCTGTCCAGCCCCCTCCCACCTGTGTCTGAATGCAAACAGGG 819

Qy     779 TGTTGCGGGGG 789
           ||   ||| |
Db     820 GGTGTTGGGCG 830
```

RESULT 9
BF594409/c
LOCUS BF594409 405 bp mRNA linear EST 12-DEC-2000
DEFINITION 7i05a05.x1 NCI_CGAP_Co16 Homo sapiens cDNA clone IMAGE:3324560 3' similar to contains Alu repetitive element;contains element MER28 repetitive element ;, mRNA sequence.
ACCESSION BF594409
VERSION BF594409.1 GI:11686733
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 405)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 404.
FEATURES Location/Qualifiers
source 1. .405
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3324560"
/clone_lib="NCI_CGAP_Co16"
/tissue_type="colon tumor, RER+"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_Co10 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1057416-1061255, and 1144584-1145351). Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 54 a 120 c 107 g 124 t
ORIGIN

```
Query Match          43.8%;  Score 405;  DB 12;  Length 405;
Best Local Similarity  100.0%;  Pred. No. 3.2e-58;
Matches  405;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     488 TGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCG 547
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     405 TGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCG 346

Qy     548 TGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTAGAGACA 607
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     345 TGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTAGAGACA 286

Qy     608 AGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAAT 667
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     285 AGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAAT 226

Qy     668 GAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTGCAGG 727
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     225 GAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTGCAGG 166

Qy     728 CACAGGGTGCCCCTGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTGTTGCGGG 787
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165 CACAGGGTGCCCCTGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTGTTGCGGG 106

Qy     788 GGCAACATGAGAGCCCCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGTGAAGAGC 847
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     105 GGCAACATGAGAGCCCCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGTGAAGAGC 46

Qy     848 CCCACCTCGGGGTCACAATAAAGTTGCCTGGTCAGGAAAAAAAAA 892
           ||||||||||||||||||||||||||||||||||||||||||||
Db      45 CCCACCTCGGGGTCACAATAAAGTTGCCTGGTCAGGAAAAAAAAA 1
```

RESULT 10
AW731946/c
LOCUS AW731946 399 bp mRNA linear EST 21-APR-2000
DEFINITION ba03e06.x1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:2823298 3', mRNA sequence.
ACCESSION AW731946
VERSION AW731946.1 GI:7632268
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 399)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library
Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml
Seq primer: -40UP from Gibco
High quality sequence stop: 381.
FEATURES Location/Qualifiers
source 1. .399
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2823298"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average

insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 53 a 116 c 108 g 122 t
ORIGIN

Query Match 43.0%; Score 397.4; DB 10; Length 399;
Best Local Similarity 99.7%; Pred. No. 5.9e-57;
Matches 398; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    493 GAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCGTGGCT 552
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    399 GAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGAAGCAGGCGAGCGTGGCT 340

Qy    553 AAGAAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTAGAGACAAGAGC 612
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    339 AAGAAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCAGCGTTAGAGACAAGAGC 280

Qy    613 CCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAGGATTCAAGAAATGAGGA 672
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    279 CCTGTGAAAAAGGCCAAAGCCACACCCAGAGCCAAGAAGCCAGGATTCAAGAAATGAGGA 220

Qy    673 GCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTGCAGGCACAG 732
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219 GCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCTGGGCTGTGCTGCAGGCACAG 160

Qy    733 GGTGCCCCTGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTGTTGCGGGGGCAA 792
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    159 GGTGCCCCTGTCCAGCCCCTCCACCTGTGTCTGAATGCAACAGGGGTGTTGCGGGGGCAA 100

Qy    793 CATGAGAGCCCCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGTGAAGAGCCCCAC 852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     99 CATGAGAGCCCCTCACCCCCAACTCTCCACTTTCAGGAGGCCCCCAGTGAAGAGCCCCAC 40

Qy    853 CTCGGGGTCACAATAAAGTTGCCTGGTCAGGAAAAAAAA 891
          ||||||||||||||||||||||||||||||||||||||||
Db     39 CTCGGGGTCACAATAAAGTTGCCTGGTCAGGAAAAAAAA 1
```

RESULT 11
BQ684424
LOCUS BQ684424 935 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT_8209062 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6260492 5', mRNA sequence.
ACCESSION BQ684424
VERSION BQ684424.1 GI:21797103
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 935)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2419 row: k column: 21
High quality sequence stop: 635.
FEATURES Location/Qualifiers
source 1. .935
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6260492"
/clone_lib="NIH_MGC_112"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 186 a 304 c 269 g 175 t 1 others
ORIGIN

Query Match 42.2%; Score 389.6; DB 14; Length 935;
Best Local Similarity 98.7%; Pred. No. 8.9e-56;
Matches 392; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy      1 CAGCCCGCTTCTCTGCCCGGAGCCATGAATCTCAGTAGCGCCAGTAGCACGGAGGAAAAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    207 CAGCCCGCTTCTCTGCCCGGAGCCATGAATCTCAGTAGCGCCAGTAGCACGGAGGAAAAG 266

Qy     61 GCAGTGACGACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGACTTGGACCTTC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    267 GCAGTGACGACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGACTTGGACCTTC 326

Qy    121 AGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCATACGATTTGCTTGGGG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    327 AGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCATACGATTTGCTTGGGG 386

Qy    181 GAGAAAGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCAGCAAACCAGGAGGAC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    387 GAGAAAGCCAAAGAGGAGATGCATCGCGTGGAGATCCTGCCCCCAGCAAACCAGGAGGAC 446

Qy    241 AAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCCCATGGTCTCC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    447 AAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCCCATGGTCTCC 506

Qy    301 ATGGTAGGAGTGCAGCTTTCTCCCCCAGTTACTTTCCAGCTCCGGGCTGGCTCAGGACCC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    507 ATGGTAGGAGTGCAGCTTTCTCCCCCAGTTACTTTCCAGCTCCGGGCTGGCTCAGGACCC 566

Qy    361 GTGTTCCTCAGTGGCCAGGAACGTTATGAAGCATCAG 397
          |||||||||||||||||||||||||||||||  |  ||
Db    567 GTGTTCCTCAGTGGCCAGGAACGTTATGANAAAAAAG 603
```

RESULT 12
BQ682257
LOCUS BQ682257 966 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT_8194971 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6259047 5', mRNA sequence.
ACCESSION BQ682257
VERSION BQ682257.1 GI:21794936
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 966)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2415 row: o column: 16
High quality sequence stop: 518.

```
FEATURES             Location/Qualifiers
     source          1. .966
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6259047"
                     /clone_lib="NIH_MGC_112"
                     /tissue_type="melanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
BASE COUNT      205 a    300 c    295 g    159 t      7 others
ORIGIN

  Query Match             41.0%;  Score 378.4;  DB 14;  Length 966;
  Best Local Similarity   74.0%;  Pred. No. 6.5e-54;
  Matches  668;  Conservative    0;  Mismatches   56;  Indels  179;  Gaps    8;

Qy       1 CAGCCCGCTTCTCTGCCCGGAGCCATGAATCTCAGTAGCGCCAGTAGCACGGAGGAAAAG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     207 CAGCCCGCTTCTCTGCCCGGAGCCATGAATCTCAGTAGCGCCAGTAGCACGGAGGAAAAG 266

Qy      61 GCAGTGACGACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGACTTGGACCTTC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     267 GCAGTGACGACCGTGCTCTGGGGCTGCGAGCTCAGTCAGGAGAGGCGGACTTGGACCTTC 326

Qy     121 AGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCATACGATTTGCTTGGGG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     327 AGACCCCAGCTGGAGGGGAAGCAGAGCTGCAGGCTGTTGCTTCATACGATTTGCTTGGGG 386

Qy     181 GAGAAAGCCAAAGAGGAGATGC-ATCGCGTGGAGATCCTGCCCCCAGCAAACCAGGAGGA 239
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     387 GAGAAAGCCAAAGAGGAGATGCNATCGCGTGGAGATCCTGCCCCCAGCAAACCAGGAGGA 446

Qy     240 CAAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCCCATGGTCTC 299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     447 CAAGAAGATGCAGCCGGTCACCATTGCCTCACTCCAGGCCTCAGTCCTCCCCATGGTCTC 506

Qy     300 CATGG--TAGGAGTGCAGCTTTCTCCCCCAGTTACTTTCCAGCTCCGGGCTGGCTCAGG- 356
           |||||    ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     507 CATGGGTAGGGAGTGCAGCTTTCTCCCCCAGTTACTTTCCAGCTCCGGGCTGGCTCAGGG 566

Qy     357 -ACCCGTGTTCCTCAGTGGCCAGGAACGTTATGAAGCATCAGACCTAACCTGGGAGGAGG 415
            ||   ||||||||||||||||||||||||
Db     567 ACCCGTGGTTCCTCAGTGGCCAGGAACGTTAT-------------------------- 598

Qy     416 AGGAGGAAGAAGAAGGGGAGGAGGAGGAAGAGGAAGAGGAAGATGATGAGGATGAGGATG 475
Db     599 ---------------------------------------------------------- 598

Qy     476 CAGATATATCTCTGGAGGAGCAAAGCCCTGTCAAACAAGTCAAAAGGCTGGTGCCCCAGA 535
Db     599 ---------------------------------------------------------- 598

Qy     536 AGCAGGCGAGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCA 595
                       ||||||||||||||||||||||||||||||||||||||||||||
Db     599 ------------------GAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAGCCA 639

Qy     596 GCGTTAGAGACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAG 655
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     640 GCGTTAGAGACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGCCAG 699

Qy     656 GATTCAAGAAATGAGGAGCCAC--GCCTTGGGGGGCACGGTGCAAAGTGGGCCTT-CCCT 712
           |||||||||||||||||||  |   | | ||||||||||||||||||||||| ||||
Db     700 GATTCAAGAAATGAGGAGCCCCCGCCTTGGGGGGGCACGGTGCAAAGTGGGCCTTCCCCT 759

Qy     713 GGGCTGTGCTGCAGGCACAGGGTGCCCCTGTCCAGCCCCTCCACCTGTGTCTGAATGCAA 772
           ||||||||||||||| |||||||||||||| |||||||  |||||| |||||||    |
```

```
Db     760 GGGCTGTGCTGCAGGGACAGGGTGCCCCTGGCCAGCCCCCTCACCTGGGTCTGAAATGCA 819

Qy     773 CAGGGGTG--TTGCGGGGGCAACATGAGAGCCCCTCACCCCCAACTCTCCACTTTCAGGA 830
           | |||| |  |   |||| |||| ||||| |||| |||||  |||||||| |||  ||
Db     820 CGGGGGGGGTTGCGGGGGNCAACCTGAGAACCCCCTCCCCCCCACTCTCCANTTTTCGGG 879

Qy     831 --GGCCCCCAGTGAAGAGCCCCACCTCGGGGTCACAATAAAGTTGCCTGGTCAGGAAAAA 888
             ||||||     || | |||| ||| ||||| | || ||   || |||| | || | |
Db     880 AGGGCCCCAGNGCAAAACCCCCCCCTTGGGGTGAAAAAAACCGTGNCTGGGCCGGGANAC 939

Qy     889 AAA 891
           |||
Db     940 AAA 942


RESULT 13
AW002370/c
LOCUS       AW002370                 361 bp    mRNA    linear   EST 09-MAR-2000
DEFINITION  wu61a07.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2524500 3',
            mRNA sequence.
ACCESSION   AW002370
VERSION     AW002370.1  GI:5849286
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 431    Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .361
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2524500"
                     /clone_lib="NCI_CGAP_GC6"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
                     from the normalized library NCI_CGAP_GC4 was prepared, and
                     ss circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     1257096-1258631, 1469064-1470983, and 1475592-1476743).
                     Subtraction by Bento Soares and M. Fatima Bonaldo.  "
BASE COUNT       50 a    107 c     93 g    111 t
ORIGIN

  Query Match             38.7%;  Score 357.8;  DB 10;  Length 361;
  Best Local Similarity   99.4%;  Pred. No. 2.5e-50;
  Matches  359;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     533 AGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAG 592
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAG 302
```

```
Qy    593 CCAGCGTTAGAGACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGC 652
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCAGCGTTAGAGACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGC 242

Qy    653 CAGGATTCAAGAAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCT 712
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CAGGATTCAAGAAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCT 182

Qy    713 GGGCTGTGCTGCAGGCACAGGGTGCCCCTGTCCAGCCCCTCCACCTGTGTCTGAATGCAA 772
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    181 GGGCTGTGCTGCAGGCACAGGGTGCCCCTGTCCAGCCCCTCCACCTGTGTTTGAATGCAA 122

Qy    773 CAGGGGTGTTGCGGGGGCAACATGAGAGCCCCTCACCCCCAACTCTCCACTTTCAGGAGG 832
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    121 CAGGGGTGTTGCGGGGGCAACATGAGAGCCCCTCACCCCCAACTTTCCACTTTCAGGAGG 62

Qy    833 CCCCCAGTGAAGAGCCCCACCTCGGGGTCACAATAAAGTTGCCTGGTCAGGAAAAAAAAA 892
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCCCCAGTGAAGAGCCCCACCTCGGGGTCACAATAAAGTTGCCTGGTCAGGAAAAAAAAA 2

Qy    893 A 893
          |
Db      1 A 1
```

RESULT 14
BF592761/c
LOCUS BF592761 361 bp mRNA linear EST 12-DEC-2000
DEFINITION 7j94b04.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:3442494 3', mRNA sequence.
ACCESSION BF592761
VERSION BF592761.1 GI:11685085
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 361)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -40UP from Gibco.
FEATURES Location/Qualifiers
source 1. .361
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3442494"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 50 a 106 c 94 g 111 t
ORIGIN

Query Match 38.7%; Score 357.8; DB 12; Length 361;
Best Local Similarity 99.4%; Pred. No. 2.5e-50;
Matches 359; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy    533 AGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAG 592
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAG 302

Qy    593 CCAGCGTTAGAGACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGC 652
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCAGCGTTAGAGACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGC 242

Qy    653 CAGGATTCAAGAAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCT 712
          ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    241 CAGGATTCAAGAAATTAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCT 182

Qy    713 GGGCTGTGCTGCAGGCACAGGGTGCCCCTGTCCAGCCCCTCCACCTGTGTCTGAATGCAA 772
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    181 GGGCTGTGCTGCAGGCACAGGGTGCCCCTGTCCAGCCCCTCCACCTGTGTTTGAATGCAA 122

Qy    773 CAGGGGTGTTGCGGGGGCAACATGAGAGCCCCTCACCCCCAACTCTCCACTTTCAGGAGG 832
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    121 CAGGGGTGTTGCGGGGGCAACATGAGAGCCCCTCACCCCCAACTCTCCACTTTCAGGAGG 62

Qy    833 CCCCCAGTGAAGAGCCCCACCTCGGGGTCACAATAAAGTTGCCTGGTCAGGAAAAAAAAA 892
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCCCCAGTGAAGAGCCCCACCTCGGGGTCACAATAAAGTTGCCTGGTCAGGAAAAAAAAA 2

Qy    893 A 893
          |
Db      1 A 1
```

RESULT 15
BF057162/c
LOCUS BF057162 361 bp mRNA linear EST 16-OCT-2000
DEFINITION 7k16g07.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:3444661 3', mRNA sequence.
ACCESSION BF057162
VERSION BF057162.1 GI:10811058
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 361)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -40UP from Gibco.
FEATURES Location/Qualifiers
source 1. .361
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3444661"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA

from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 50 a 107 c 93 g 111 t
ORIGIN

Query Match 38.7%; Score 357.8; DB 12; Length 361;
Best Local Similarity 99.4%; Pred. No. 2.5e-50;
Matches 359; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy     533 AGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAG 592
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AGAAGCAGGCGAGCGTGGCTAAGAAAAAAAAGCTGGAAAAAGAAGAAGAGGAAATAAGAG 302

Qy     593 CCAGCGTTAGAGACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGC 652
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CCAGCGTTAGAGACAAGAGCCCTGTGAAAAAGGCCAAAGCCACAGCCAGAGCCAAGAAGC 242

Qy     653 CAGGATTCAAGAAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCT 712
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CAGGATTCAAGAAATGAGGAGCCACGCCTTGGGGGGCACGGTGCAAAGTGGGCCTTCCCT 182

Qy     713 GGGCTGTGCTGCAGGCACAGGGTGCCCCTGTCCAGCCCCTCCACCTGTGTCTGAATGCAA 772
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     181 GGGCTGTGCTGCAGGCACAGGGTGCCCCTGTCCAGCCCCTCCACCTGTGTTTGAATGCAA 122

Qy     773 CAGGGGTGTTGCGGGGGCAACATGAGAGCCCCTCACCCCCAACTCTCCACTTTCAGGAGG 832
           |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db     121 CAGGGGTGTTGCGGGGGCAACATGAGAGCCCCTCACCCCCAACTTTCCACTTTCAGGAGG 62

Qy     833 CCCCCAGTGAAGAGCCCCACCTCGGGGTCACAATAAAGTTGCCTGGTCAGGAAAAAAAAA 892
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CCCCCAGTGAAGAGCCCCACCTCGGGGTCACAATAAAGTTGCCTGGTCAGGAAAAAAAAA 2

Qy     893 A 893
           |
Db       1 A 1
```

Search completed: January 24, 2003, 18:43:01
Job time : 2248 secs